NEWCASTLE UNIVERSITY

THE EFFECT OF EXPLICIT INSTRUCTION AND AUDITORY/AUDIO-VISUAL TRAINING ON CHINESE EFL LEARNERS' PERCEPTION OF INTONATION

A THESIS SUBMITTED TO

THE FACULTY OF HUMANITIES AND SOCIAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY (INTEGRATED)

SCHOOL OF EDUCATION, COMMUNICATION AND LANGUAGE SCIENCES

BY

YILING CHEN

NEWCASTLE UPON TYNE, UNITED KINGDOM

JANUARY, 2020

Abstract

Intonation accounts for a big part in speech intelligibility and is notoriously difficult to be acquired by L2 learners. The bulk of research on L2 intonation has focussed on the examination of learners' intonational performance at the phonetic and phonological levels using perceptual and/or production tasks; however, empirical studies on whether and how intonation training can help improve learners' performance are surprisingly scarce. This study fills this gap by devising instruction and training materials which were meticulously tailored for Chinese learners of English, the largest population of English learners in the world. The participants were 60 English-related majoring students from Newcastle University, most of whom wanted to become English teachers following their studies. They were pseudo-randomly mapped into three groups according to their overall English proficiency. Two of the groups were taught explicitly on the forms and functions of English intonation but one self-practiced auditorily on *Audacity* whereas the other audio-visually on *Praat*. The third group, which served as control, did not get any intonation training. Learners' competence of intonation was assessed by a comprehension task before, immediately after, and two months after the three-week training course. Ten native speakers of Southern British English were recruited for the pre- and post-test to set a baseline for the analysis of learners' performance. The results are: 1. Chinese EFL learners did significantly worse than native speakers in terms of understanding intonation meanings contrasted by accentuation, phrasing, and tone. 2. Learners' comprehension ability was improved immediately after the training for all three aspects. 3. The training effect remain in the delayed post-test. 4. The audio-visual group did not perform significantly better than the auditory group. The results indicate that certain aspects of intonation are teachable and learnable, and tailor-made instruction and materials are effective and applicable in use. This study provides English teachers in China with novel ways to equip Chinese EFL learners with greater intonational competence.

Acknowledgements

I still cannot believe that I can finish this thesis and finally touch upon this bit. First of all, I would love to thank myself who did not give up on writing-up after experiencing so many unexpected and horrible things during the last year of this PhD. Secondly, I feel so lucky for meeting Jack An, my husband, at the beginning of my PhD life, otherwise it would have lost its richness and laughters. I am also indebted to my parents who have been supporting me emotionally and finiancially. Without them, I would not even have had a chance to pursue a doctoral degree.

Now my thanks have to go to my supervisors, Dr. Ghada Khattab and Dr. Jalal Al-Tamimi, who have alway been there for my academic needs. Without their support, I cannot reach to this end. I would love to give my special thanks to Bodo Winter, who helped me on the GLMM modelling when he was visiting Newcastle University and offering us a fantastic workshop on statistics with advanced methods. I also want to thank my colleagues who have worked with me, accompanied me both in and out of the office. They are : Pachonak, Kittikana, Cecilia Devers, Katerina Nimz, Oksana Lyalka, Dan McCarney, Hajar Moussa, and Maha Jamin. Our friendship will be cherished for the rest of my life.

Finally, I have to thank my participants who genuinely trusted me by giving their precious time on my experiment. I was truly encouraged by their unconditional support. This thesis would not come out without their contribution.

To my grandma

Table of Contents

List of Tables

List of Figures

Chapter I: Introduction

This chapter begins with an overview of the research background for the present study, followed by a clarification of some important conceptual terminologies that pervade the whole thesis. Then an outline of this thesis will close this chapter.

1.1 Background

The past century has witnessed an evolvement of ESL (English as a Second Language) or EFL (English as a Foreign Language) pronunciation teaching from an "imitative-intuitive"[1] approach to a more analytically-oriented approach (roughly from 1900s to 1970s) since the advent of the International Phonetic Alphabet (IPA) (Murphy and Baker, 2015). Before 20th century, the teaching of pronunciation was deemed "largely irrelevant" (Celce-Murcia *et al.*, 2010: 3). During the era of analytic orientation to the instruction of pronunciation, audiolingualism was probably one of the best-known approaches whereby accurate pronunciation of individual sounds was expected from students (Pennington and Richards, 1986) through a considerable amount of repetition drills that were mainly restricted to segmental phonemes (Lado, 1957). Suprasegmental features such as stress, rhythm and intonation were rarely instructed (Munro and Derwing, 2015) even though they were also scheduled in a typical audiolingual classroom at that time (Morley, 1991). Exercises used for audiolingual teaching were normally de-contextualised, in that students might be left with no clue what they were saying and why and when a particular intonation pattern should be used (Murphy and Baker, 2015), as words, phrases, and sentences in such exercises were merely repeated in isolation (e.g. Nilsen and Nilsen, 1971).

[1] This term was firstly used by Marianne Celce-Murcia and her colleagues in their book *Teaching Pronunciation: A Course Book and Reference Guide* (2010). It means that the teaching of pronunciation is solely by means of listening and imitating without explicit instruction on the phonetic and phonological details.

With the fall of audiolingualism, pronunciation teaching was marginalised or even avoided (Celce-Murcia *et al.*, 1996) as people were inclined to believe that it was unrealistic for adult L2 learners to achieve a native-like accent. Only until 1980s, when the Communicative Language Teaching (CLT) framework started to dominate the L2 teaching, has pronunciation teaching seen a resurgence under the expanded impact of CLT (Murphy and Baker, 2015). Since then, learners' communicative competence has begun to be considered as the ultimate goal of pronunciation teaching, given the belief that "without communicative intent, pronunciation is not true speech; it is no more than the manipulation of linguistic forms" (Prator and Robinett, 1985: xvi). A growing number of empirical research studies were conducted to explore the aspects of pronunciation that contribute to learners' speech intelligibility (Derwing *et al.*, 1997; Tajima *et al.*, 1997; Munro and Derwing, 1998). Specialist writers and material developers tended to direct their attention to suprasegmentals that were thought to be of more communicative values (Lightbown and Spada, 2006; Celce-Murcia *et al.*, 2010).

Although the current trend of pronunciation and intonation teaching is intelligibility-oriented, English learners seem to favour native accent (Remiszewski, 2005), which is discouraged in Jenkins's Lingua Franca Core (LFC) (2000, 2007). LFC is a pronunciation teaching paradigm which includes only the pronunciation details that foreign language learners have to acquire for intelligible communication both among native and foreign learners of English and among learners with different L1 backgrounds. In fact, language learners tend to be sensitive to foreign accent (Derwing and Munro, 2015). Some have even been documented having experience of being discriminated (e.g. Munro, 2003). It is not wise to ignore learners' needs for the learning of intonation, as stated by Jenkins (2000: 101): "It's important not to patronise those learners who wish to work towards the goal of a NS (native speaker) accent by telling them that have no need to do so".

L1-Chinese learners of English, as the largest population of ESL/EFL learners in the world and as the research target in the current project, have long been at the centre of discussion in

L2 pronunciation research, but the research in the teaching and learning of English pronunciation for this particular group is limited. The following paragraphs attempt to delve into the reasons behind the scene.

In the wave of the higher English education reform in mainland China since 2000,[2] the CLT framework saw its first official recognition in the *College English Curriculum Requirements* (CECR) (2007) published by the Ministry of Education.[3] In CECR, the pronunciation teaching involved a balanced curriculum for segments and suprasegmentals, and it was reflected in the *Standards for Oral Proficiency in English* (SOPE-RUC), an assessment scheme for Chinese college students' oral English proficiency, established by the Foreign Language Research Committee of Renmin University (Gu *et al.*, 2013). In line with the Common European Framework of reference for languages (CEFR) (Council of Europe, 2001), SOPE-RUC classifies six levels of English proficiency from the least to the most advanced levels (A1, A2, B1, B2, C1, C2). In each of these levels, suprasegmentals weigh equally with the individual sounds. At the most advanced C2 level, learners are expected to perceive and produce appropriate stress, intonation, and rhythm to communicate smoothly without any difficulty. To be more specific, apart from the precise articulation of vowels and consonants, learners will be able to utter semantically coherent English by phrasing and stressing their utterances at proper positions, to manipulate phonological processes such as linking, assimilation, and coarticulation to make their speech sound naturally fluent. The only criterion that distinguishes C1 from C2 is that the minor errors of the production of phrasing and stressing are allowed in C1 as long as the meaning of their speech is not compromised (Gu *et al.*, 2013: 9).

Unfortunately, the delivery of such a curriculum seemed to be far from satisfactory based on a national investigation of the state of English teaching from 530 universities (see Wang and Wang, 2011). One of the key reasons was the CECR prioritized the teaching of reading skills

[2] This excludes Hong Kong and Macao Special Administrative Regions where have their own independent ministry of education.
[3] English is a compulsory subject throughout the secondary and tertiary levels of education in mainland China.

while the listening and speaking were just in secondary place. The pronunciation teaching in vast majority of the universities in China (except the English majors) was treated as a trivial part of the general English course in which teachers would normally choose to ignore it and spend most of the time on vocabulary and grammar that are essential for reading. As a natural consequence, oral proficiency was rarely tested (Wang and Wang, 2011). Even for the English majors to whom pronunciation is taught as an independent course, the teaching of suprasegmentals has been mostly ignored partially due to the ingrained belief that a mastery of segmental pronunciation would lead to clear speech. But even if some Chinese speakers pronounce every English word clearly, they were still difficult to be understood (Song and Lan, 2010).

With an increasing demand for graduates who are communicatively competitive in a more globalised society, the National Foreign Language Teaching Advisory Board has proposed the *Guidelines on College English Teaching* (GCET) in 2016, to replace the CECR as the new nationwide principles of the English teaching at tertiary level. In this new guideline, the teaching of listening and speaking have surpassed reading and become the top priorities (Wang, 2016). Other modifications include the transformation of the classroom setting from teacher-centred to teacher-directed and student-centred, the inclusion of more authentic and culturally associated linguistic contents, and the incorporation of trending techniques (e.g. computer programs and online open sources) to the class, etc. Furthermore, teacher training has been particularly brought up to date for healthy and sustainable development in higher education. An English teacher with "high-quality" skills should be able to keep up with the latest research achievements and their pedagogical implications and subsequently apply them to teaching (Wang, 2016; Jia, 2017). The GCET will be put in use as soon as it is approved by the Ministry of Education.

English language teaching, with these new changes, should in principle lead to better focus on oral and aural skills, but it is confronted with severe challenges that communication-oriented pedagogy would hardly dominate English classrooms at least in the short run. The teaching of

listening and speaking before college would be extravagant hopes within China's once-in-a-lifetime college entrance examination system (*Gaokao*), as both teachers and students will spend most of their time on reading and writing that are more crucial for the *Gaokao*. It means that the majority of Chinese learners will just begin to get specific instruction on speaking with very limited time on pronunciation once they start their college life (Gu *et al.*, 2013). Speaking or pronunciation teachers in higher education, however, have little knowledge of English phonetics and phonology (Shi, 2010), so their conception of a pronunciation course might still be restricted to the way they used to get trained when they were at schools. As G. Wang (2010) portrays, a typical English pronunciation class at college in China is still following the traditional way of "pronunciation correction" which is rigidly audiolingualism-based. That means native-like pronunciation of individual words is the main goal while suprasegmental features are usually ignored (Tian, 2010), leaving the adoption of the assessment scheme mentioned above highly questionable.

The following chart summaries the relevant contents included in most widely used pronunciation textbooks in colleges in China.

Table 1. 1: Summary of the English pronunciation textbooks used in Chinese universities

Title (Authors)	**Target students**	**Pages on suprasegmentals (percentage)**	**Pages on intonation (percentage)**	**Featured intonation knowledge and practice**
A Concise Guide to Contemporary English Pronunciation (Zhu, 2003)	General college	30 (24.6%)	8 (6.6%)	Seven pitch contours symbolized by tadpoles within 5-level musical scale, including high fall, low fall, low rise, high rise, level, fall-rise, rise-fall with few sentences as examples, accompanied by basic pragmatic and emotional meanings
English Pronunciation & Intonation for Communication: A Course for Chinese EFL learners (Wang, 2005)	**English major**	88 (48.4%)	37 (20.3%)	Intonation unit (pre-head, head, nucleus, and tail), nucleus (the most stressed syllables) and tone (falling, rising, and falling-rising) Emphasis on intonation forms with a brief introduction on intonation functions (attitudinal, accentual, grammatical, and discourse function)
A Handbook of English Pronunciation Exercises for First-Year College Students (Zhang, 2006)	General college	46 (34.8%)	26 (19.7%)	Sentence contours (simple [rising vs. falling] vs. combined [falling-falling, rising-falling, falling-rising]) The intonation of question tags, vocatives, parentheses, reporting phrases Sentence contours only
English Pronunciation and intonation: Practice (Wang, 2011)	**English major**	136 (70.1%)	43 (22.2%)	Intonation unit (pre-head, head, nucleus, and tail), nucleus (the most stressed syllables) and tone (falling, rising, and falling-rising) with a brief introduction on intonation functions (accentual, grammatical, discourse, and attitudinal). Emphasis on intonation forms with a brief introduction on intonation functions An abundance of exercise in the form of listening and reading tasks

English Pronunciation for College Students: A Fundamental and Advanced Course (Wang *et al.*, 2011b)	General college	84 (29.7%)	42 (14.8%)	Intonation unit (head, nucleus and tail), tone types exemplified by attitudes and sentence types. Tone types include falling (high fall, medium fall, low fall, rise-fall) rising (high rise, medium rise, low rise, level, and fall-rise), Register and key as a function of emotional expression Practice utterances given with pitch contours together with tadpole symbols
A practical Course for English Pronunciation Skills (Yang *et al.*, 2013)	General college	68 (25.4%)	19 (7.1%)	Sense group (cutting rules according to syntax), five tone types including falling, rising, rising-falling, falling-rising, and level. The use of tone types are associated with sentence types, such as falling tones for declarative sentences, imperative sentences and exclamations, etc.

It is not surprisingly noticed that the two textbooks in which suprasegmentals account for the greatest proportions (in bold) are specifically designed for students in English majors, while the other four for general college students still emphasize the teaching of segments as the pages on suprasegmentals are less than 35%. Intonation for most of these textbooks is no more than 20%, indicating that even for English majors, intonation seems to be treated as a peripheral aspect by material writers. A more serious issue concerns with the outdated description of intonation forms and functions and an overgeneralisation of the relationship between intonation patterns and sentence types or between patterns and attitudes, which entirely echo the problems of the teaching materials back in 90s in North American as reviewed by Levis (1999). Moreover, none of these textbooks draws on empirical research on learners' acquisition of English intonation, not mentioning taking learners' difficulty into account, though they are all claimed to be specialised for Chinese learners. For instance, Wang (2005, 2011) overemphasizes the phonetic manipulation of durational lengthening and intensity amplifying for accentuated words which have proved to be of little difficulty for Chinese learners of English (Chen, 2008a; Wang *et al.*, 2011a).

All in all, pronunciation teaching in China has reached a turning point. In light of the new guidelines, a considerable amount of relevant research, updated teaching materials, and advanced pronunciation teachers specified for Chinese learners of English are in urgent need.

Apart from the reasons stated above, there have been some longstanding unresolved issues from a theoretical and pedagogical perspective that seem to impede the application of these guidelines to practice. The first and foremost is that research in the learning and teaching of pronunciation is lagging behind within the domain of L2 acquisition and applied linguistics, as pronunciation has not been a priority for most L2 researchers. Deng et al. (2009) reviewed 14 influential journals in the field of applied linguistics and L2 acquisition during the period of 1999 to 2008 and found that none of these journals had published more than 10 percent of its articles on pronunciation, not to mention suprasegmentals.

As reviewed before, it has long been recognised that pronunciation and intelligibility are important for L2 teaching since the introduction of CLT (Munro and Derwing, 1995; Jenkins, 2000; Jenkins, 2004), and the question has been converted from "whether" to "what" and "how" to teach pronunciation (Morley, 1991). This conversion has given rise to a number of empirical studies on L2 speech *intelligibility* and *comprehensibility* (see the detailed definition of both terms in section 1.3), which endeavour to pinpoint the pronunciation aspects that influence native listeners' comprehension of learners' speech and the degree of difficulty for them to understand non-native utterances. Suprasegmental features have been found to contribute greatly to the understanding of learners' speech. For example, timing factors (duration of individual sounds, words and sentences) can significantly affect the intelligibility of L2 speech (Busà, 2008), and the faster L2 learners speak English, the easier a native English speaker will understand what they said as it is more native-like (Munro and Derwing, 1998; Derwing and Munro, 2001; Munro and Derwing, 2001). Tajima et al. (1997) even found that the intelligibility of Chinese learners' English was significantly improved after a temporal modification to native-like values, and the intelligibility of native speakers' was apparently decreased with a temporal modification to Chinese-like. Misplacement of word stress was also reported to be detrimental to the intelligibility (Lepage and Busà, 2014; Chen, 2015a). The pivotal role that intonation structure plays in the communicative functions of spoken discourse has also been identified (Derwing *et al.*, 1997). All these findings mentioned above shed light on "what" to teach in a pronunciation class.

Another gap lying in between the translation of empirical research into pedagogy might be the uncertainty of the learnability of suprasegmentals and the effectiveness of teaching suprasegmentals, as some applied linguists might believe that suprasegmental features are used subconsciously such that L2 learners might not be able to manipulate them voluntarily (Setter and Jenkins, 2005: 2). Most of the Chinese learners of English, unfortunately, fall into this category. This is why, to the author's knowledge, an intervention study is needed, because it is very unlikely for learners themselves to be aware of suprasegmentals like they

are of segments that are represented in orthography (Derwing and Rossiter, 2002). One can not deny the fact that "the empirical research on pronunciation instruction is the least developed line of work among L2 pronunciation research" (Derwing and Munro, 2015: 4). The number of papers on suprasegmental instruction is even smaller (Thomson and Derwing, 2014), but the existing ones have opened up a door to a deeper investigation for the worthiness of suprasegmental teaching. Zhou et al. (2012), for instance, examined the production of four English tones (high fall, high rise, fall-rise, and complex level) by 31 English-majoring Chinese college students before and after 18-hour explicit instruction on the forms and functions of these tones, and found that learners' phonetic performance of all these tones had improved. Training on the global intonation contours of yes/no questions and wh-questions was also found significantly effective for an improvement of learners' performance (Le and Brook, 2011). An intervention specified for intonation functions was shown to be helpful for Mandarin speakers' production of overall pitch contours (Gao, 2015). Studies comparing the effectiveness of segmental instruction to that of suprasegmental instruction have revealed that only with training on suprasegmentals was learners' extemporaneous speech perceived with greater comprehensibility (Derwing *et al.*, 1998; Derwing and Rossiter, 2003). Elliott's (1997) finding was consistent with Derwing and her colleagues' in which segmental instruction did not significantly improve spontaneous utterances produced by English learners of Spanish. The above-discussed empirical findings show that suprasegmentals are (at least partially) learnable and teachable, and are worth being prioritised in pronunciation teaching.

L2 learning and teaching is not the only area in which prosody has been neglected; research on the L2 acquisition of prosody seems far from extensive compared to segments (Gut, 2009; MacDonald, 2011). Theories of second language speech acquisition such as the Perceptual Assimilation Model (Best, 1995) and the Speech Learning Model (Flege, 1995) are confined to segments. Although prosodic errors that L2 learners make at different phonetic and phonological dimensions have been unravelled by dedicated researchers (e.g. Anderson-Hsieh

and Venkatagiri, 1994; Wennerstorm, 1994; Altmann, 2006; Trofimovich and Baker, 2006; Trouvain and Gut, 2007; Chen, 2008a; He *et al.*, 2010; O'Brien and Gut, 2010; He *et al.*, 2012b), there has been no indication that all these discovered errors need pedagogical intervention, otherwise the underlying principle for teaching would be nativeness-oriented rather than intelligibility-oriented. Because most of these errors were defined by significant deviations from the native norms rather than perceptual difficulty that native listeners have to access the intelligibility of L2 learners' speech (Derwing and Munro, 2015).

Last but not least, technology-infused pronunciation teaching has been a heated topic in related fields of applied linguistics; there is particular interest in software programs and applications that can visualise both native speakers' and learners' speech on the computer screen. Proponents for using these visualisation technologies claim that learners' perceptual bias of phonetic details can be mitigated through exposure to the visualisation of these details, and thus result in an improvement of productions (Wilson, 2008; Gorjian *et al.*, 2013; Olson, 2014). There is no doubt that innovation in advanced technologies "has greatly expanded the possibilities for pronunciation instruction", as recognised by Derwing and Munro (2015: 23). Policy makers in China have also encouraged adoption of a computer-assisted pedagogical approach as reviewed for the *Guidelines on College English Teaching* (2016). However, its effectiveness regarding communication enhancement is yet to be known (Busà, 2008; Derwing and Munro, 2015). Some researchers even warned that over-reliance on these technologies might cause counter effects, as the visual feedback can be misleading sometimes (Chun, 2013). Learners' errors shown by the visual feedback may be just a reflection of environmental effects, e.g. background noises (Godwin-Jones, 2009). Nevertheless, L2 learners tend to be more active and motivated, less anxious and more relaxed in a classroom equipped with technologies (Gorjian *et al.*, 2013), so a theoretically and pedagogically motivated investigation of technology-assisted intervention is worthwhile, as it bears potential contributions to the understanding of L2 acquisition of prosody and its possibility in the application of empirically-evidenced pedagogy.

1.2 Summary

By reviewing the development and current status of pronunciation teaching worldwide and domestically in China, the present study is motivated and driven by filling the research gaps that are identified as the following. First, the intelligibility principle has been widely acknowledged as the essence of pronunciation teaching; prosodic aspects seem to be encoded with a great deal of *intelligibility* and *comprehensibility* (see the definitions of both terms in section 1.3), but many English instructors in China today still tend to rely on the reduction of foreign accent by emphasizing individual segments (Wang, 2010). This is partly due to their lack of awareness of the importance of prosody (Zhang, 2004; Wang, 2013), and/or because they have a poor command of prosody or prosodic knowledge themselves, which leads to a lack of confidence in teaching it (Chun, 2002; Shi, 2010). Second, instructional materials grounded on empirical research for a particular group of L2 learners (i.e. Chinese) are rare, so are the implications from the controlled experimentation of prosodic intervention. Third, how prosodic features are acquired and whether computer-assisted pedagogy facilitates acquisition have unsatisfactorily addressed. Teachers are normally blamed for the failure of learners' acquisition of prosody; they are encouraged to catch up with up-to-date L2 research and adjust the way they teach while at the same time having to improve their own pronunciation (GCET, 2016, Setter and Jenkins, 2005; Shi, 2010). It sounds fair to ask teachers to work on improving their own production, but it seems unfair to require every instructor to "re-invent the wheel for pronunciation teaching by finding effective techniques through trial and error" (Derwing and Munro, 2015: 25). It is for the most part the researchers' duty to raise teachers' awareness of the priority in pronunciation teaching, to present solid evidence of what aspects of pronunciation are worth teaching and what parts of learners' errors need intervention, and to construct teaching materials and technology-assisted activities that are empirically supported. Only by doing so can teachers use their limited time on pronunciation with an approach that would benefit the students most. The current research aims to do this.

1.3 Terminology

An array of L2 research concerned with intelligibility and comprehensibility has been spawned since the advent of CLT. Among those pioneering works done in 1990s, Tracey Derwing and Murray Munro are perhaps the most distinguished specialists who have clarified the definition of these terms. Therefore, the relevant notions used in the current study follow the definitions in their latest published book *Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research* (Derwing and Munro, 2015).

> *Intelligibility*—the degree of match between a speaker's intended message and the listener's comprehension;
>
> *Comprehensibility*—the ease or difficulty a listener experiences in understanding an utterance;
>
> *Accent*—a particular pattern of pronunciation that is perceived to distinguish members of different speech communities.[1]

The definitions of suprasegmentals, prosody and intonation however, are far from unanimous among theoretical and applied researchers. According to Cutler et al. (1997) and Cutler and Swinney (1987), *prosody* is a term used interchangeably with *suprasegmentals* and both refer to a series of features larger than segments, including lexical stress, phrase and compound stress, sentence stress, rhythm and intonation. These prosodic or suprasegmental features are phonetically realised by fundamental frequency, duration and intensity. Crystal (2008) defines prosody as subordinated to suprasegmentals, which pertain to both linguistics and paralinguistics, while prosody is the term favoured in the linguistic end.

As for the definition of *intonation*, the inconsistency falls on the phonetic dimensions that theoretical researchers have used to describe intonational phonology, but no one would doubt that fundamental frequency (henceforth F0) plays the most important part (Roach, 2009). In a

[1] "Accent" used in this paper is restricted to the pronunciation rather than vocabulary or grammatical differences that distinguish a speech community from anther.

narrow sense, intonation refers to F0 patterns that "convey information beyond lexical meanings" (Xu, 2017: 458). In the *Longman Dictionary of Language and Applied Linguistics* (Richards and Schmidt, 2002: 272), intonation is a collection of "phenomena" in speaking, involving raising and lowering the pitch of people's voice, forming pitch patterns, increasing the duration and loudness of some syllables in the utterances, and changing the speech rhythm. In this broader sense, *intonation* is loosely equivalent to *prosody*. This thesis will adopt the broader definition of intonation for English, as it will not study intonation of pitch variations only as if they were isolated from the rhythmic system (Nolan, 2006), and in fact, salient pitch changes "always co-occur with syllable lengthening, which is usually considered part of the stress/rhythm system" (Levis, 2005: 349). For English teachers, it would be difficult to illustrate the use of a particular pitch contour without mentioning the syllable(s) that this contour attaches to. Wells (2006) also enunciated the importance of learning how stressed and unstressed syllables alternate as part of the linguistic functions that pitch patterns are encrypted for on the first page of his book *English Intonation: An Introduction*, a textbook for researchers, teachers and English learners.

1.4 Outline of the thesis

The literature review chapter (Ch.2) focuses on an elaboration of English intonational system with a comparison to that of Mandarin Chinese, followed by a comprehensive review of previous research on L2 learners' acquisition of English intonation/prosody, primarily targeting Chinese ESL/EFL learners. Following next is an overview of empirical studies on the effectiveness of technology-assisted prosodic training which will close the chapter of literature review. The methodology chapter (Ch.3) starts with the research questions inspired by the literature review in Chapter 2, followed by a detailed description of the participants in this study. It also presents the construction of the training materials for the instruction and practice enlightened by the reviews in Ch.2. Then an introduction of the comprehension experiment will close this chapter. Chapter 4 presents the results of the experiment. Chapter 5

discusses the major results with references to the literature. Chapter 6 comprises the concluding remarks of this thesis and its contribution to the teaching and learning of English pronunciation in China and its suggestions for the future study.

Chapter II: Literature Review

2.1 English intonation

The analysis of English intonation started to thrive in the first half of the twentieth century on both sides of the Atlantic and seemed to develop into its own traditions: one is the configuration approach that dominates intonation analyses across the Britain, also known as the British school; the other is the level or phonemic approach that originated from the United States. Both approaches will be illustrated in detail with a brief review of their historical development. Other intonation models such as the IPO model ('t Hart *et al.*, 1990), the PENTA model (Xu, 2004a; Xu, 2005), etc. have also established their fame among researchers who are perception-inspired and function-driven respectively. Since they are beyond the scope of the current study, they will not be reviewed in this chapter.

2.1.1 British approaches

In Britain, most scholars working in the field of English intonation have relied heavily on previous studies that gradually form the British tradition: the configuration approach to intonation (Zhang, 2008). Palmer (1922) initiated the concept of tone unit, tone group, or sense group, the basic structure of English intonation, in his book: *English Intonation with Systematic Exercises*. The overall shape of the pitch contour on the basis of a tone group was analysed via auditory perception. The internal structure of a tone unit was then further developed by Kingdon (1958) in his work *The Groundwork of English Intonation*, and O'Connor and Arnold's (1961) *Intonation of Colloquial English* that has been consistently used in later analyses of British English. Each tone group or unit has a tune pattern that can be broken up into smaller components: a pre-head, a head, a nucleus, and a tail. The head starts from the first stressed syllable (the onset) and expands to the syllable right before the nucleus, the most salient syllable in a unit. Any unstressed syllables before the head is called the pre-

head, while those after the nucleus is termed the tail. Only the nucleus is obligatory, while the other three are optional. O'Connor and Arnold (1961) made an attempt to further distinguish phonetic and phonological levels of analysis for intonation. They categorized thirteen phonetically contrastive tunes: two pre-heads, four heads, and seven nuclei into ten phonologically distinctive tune combinations based on the attitudes they convey (three redundant tune combinations omitted). Although their book provides a full enough account of attitudinal variables that exist in English (Crystal, 1969), the tune groupings show considerable redundancy, and it is not easy to find a one-to-one match between tone patterns and attitudes (Chun, 2002; Zhang, 2008).

Later on, in contrast to O'Connor and Arnold (1961), Halliday (1967, 1970) established phonological contrasts of intonation on the basis of meaningful grammatical units which have gone far beyond what previous linguists referred to as sentence types (e.g. statement, question, command, etc.). Halliday's grammatical distinctions encompass information distribution or focus, sentence structure, clause structure, sentence function, reservation, agreement, commitment, involvement, force, negation type, request type, vocative function and co-ordination contrast, all of which could be realised by three different kinds of choices that are coined with the same initial "T": Tonality, Tonicity, and Tone. Tonality means the division of an utterance into tone groups or intonation phrases each of which concerns a single piece of information that the speaker intends to convey. In most cases tonality is associated with the grammatical units, but there is no absolute correspondence (Wells, 2006). It is speakers' own choices of where to signal a new start of a tone group, but there are some shared rules in English governing the chunking of spoken discourse, for instance, a final reporting phrase normally not being an independent tone group (Cruttenden, 1997), e.g. //"What are they?" she asked him at last.// (extracted from Chen, 2008a: 9).[1] The whole utterance tends to be treated as one tone group, as the final reporting phrase itself is not carrying important information that deserves the listener's attention. Cruttenden (1997)

[1] "//" indicates the boundary of a tone group. The final reporting phrase is underlined.

proposes four external criteria for the identification of tone group boundaries that are phonetically indexed, i.e. 1. pause, including filled pause (/m/, /ə/ when thinking or hesitating, mostly in spontaneous speech), silence, and breath, 2. anacrusis (weakened production of the pre-head), 3. final syllable lengthening, and 4. pitch reset. But all four phonetic cues bear an inherent problem that they can either signal tone group boundaries, or mark hesitations (ibid.: 29-34). As a result, two internal criteria have been put forward to echo the basic structure of a tone group: 1) there must be at least one stressed syllable within a tone group; 2) there must be a pitch change on the stressed syllable. If they don't show up, even though the four external cues are all present, the tone group does not exist.

Tonicity refers to the placement of tonic syllable, or nucleus. It is also known as sentence stress/accent, intonation centre, etc. As mentioned before, the major part of the pitch movement is hooked on the nucleus which carries the information focus. Normally, words that carry new or unpredictable information in an utterance gain more prominence than the given/old, or contextually hinted information (Celik, 2001). In English, the unmarked position of the nuclear syllable is almost always on the final content word's stressed syllable of a tone group (Tench, 1996; Cruttenden, 1997; Wells, 2006), because, as Halliday (2004) explains previously, mentioned information is likely to precedes the new. But this is not always the case. New information can also come with function words or non-final content words where they become the marked positions of the nuclei. Emphatic and contrastive information can be accented on marked positions as well. The most consistently researched phonetic dimensions involved in a nuclear syllable include higher intensity, longer duration, more expanded F0 range, and more precise articulation of vowel than that of non-nuclear syllables (Sluijter and van Heuven, 1996; Pan *et al.*, 2005; Gut *et al.*, 2013).[2]

Tone by definition is the dramatic pitch movement on the nuclear syllable and its following tail syllables if there is any. In this sense, tone is more frequently termed as nuclear tone so as to distinguish the pitch patterns of the head and the pre-head. In British English, five nuclear

[2] Includes non-stressed syllables and stressed but non-accented syllables.

tones are often discussed. They are falling, rising, falling-rising, rising-falling, and level, in which falling, rising, and falling-rising are the very basic nuclear contours distinctive in functions that L2 learners should pay special attention to (Wells, 2006). Halliday (1967) also proposed five tonic contours that are slightly different from the above, but they are rarely used nowadays (Levis and Wichmann, 2015). Literally, a falling tone is a downward movement from a relatively high pitch to a lower pitch. There is normally a step-up in pitch preceding the nuclear falling tone, but they are "irrelevant in determining the nuclear tone" (Wells, 2006: 18). If the nuclear syllable is located at the final position of a particular intonation phrase, the downward movement would complete within this syllable. But if the nuclear syllable is followed by a tail (a string of unstressed syllables), the fall would normally finish within the nuclear syllable and stay at its lower level across the tail and reach its lowest point at the final unstressed syllable. The rising tone, on the contrary, begins low and ends high, and if it is hooked on the nuclear syllable at the sentence final position, the rising process would be accomplished within this syllable, whereas those on a non-final position rising tones would be gradually heightened up until the end of the utterance. Falling-rising tone is a complex tone that is frequently used in the Received pronunciation (RP) and General American (GA) while less likely in other varieties of English (Wells, 2006). The pitch movement goes downwards first then upwards. The highest point usually appears at the beginning (anywhere between the mid and high pitch range of the speaker) and reaches to its trough and finally rises to a middle point. The whole process finishes on the nuclear syllables at the final position of the intonation phrase. If any unstressed syllables follow the nucleus, the falling part takes place on the nucleus or between the nuclear and its next syllable; the rising part covers the rest syllables of the tail (Wells, 2006: 23)

The transcription of intonation in the British school is normally integrated into the text although some symbolic variants have been found. Tonality is marked by vertical bars between successive tone groups, i.e. | or ||, or in italic /, //. Tonicity is either underlined or capitalised or altogether. Tone is usually placed immediately before the nuclear syllable by

either sloping arrows ↗, ↘↗, or strokes \, /, ᵛ (Halliday, 1970; Brazil *et al.*, 1980; Cruttenden, 1997; Wells, 2006).

2.1.2 American approaches

In American approaches, intonation is considered as a sequence of discrete pitch levels; thus the minimal unit of intonation is pitch phonemes rather than pitch contour. The representatives of this approach are Pike (1945) and Trager and Smith (1951). The American school argued that the description of pitch contours such as falling, rising, or rising-falling was not enough unless the internal structures of these contours were specified with respect to the variations within the same type of contour. Inspired by the structural linguists' phonemic theory, Pike (1945) distinguished four pitch levels numbered from 1 to 4 as phonologically contrastive phonemes: 1=extra high, 2=high, 3=mid, 4=low. These pitch heights are relatively defined on the basis of one's voice range, resulting in the variations from individual to individual. Intonation contours (or tunes), from Pike's standpoint, are composed of the interpolations between the levels that are assigned to the syllables. It is the contour that make sense while these phonemic pitch levels are meaningless on their own. The tone level marks the starting points, the direction changing points, and the ending points of the contours. The primary contour (the nucleus in the British school) is symbolised with °. The following example shows the typical phonemic transcription of intonation (extracted from Levis and Wichmann 2015:140):

I want to go home

3– 2°– 4

This four-levelled system generated as many contours as Pike attempted to analyse according to the attitude that each of these contours distinguished from each other. This is possibly where Pike's system was criticized most, as many of the contours he interpreted as communicatively discrete were actually just shown as a matter of degree difference (Levis

and Wichmann, 2015). Another criticism concerns the phonemic level itself, as Cruttenden (1997: 38-39) doubted the uniqueness of this four-level pitch phoneme; it might be able to have three or five levels that could work as well as four levels.

Pike's phonemic approach to treating intonation contours as distinct pitch levels was further developed by Janet Pierrehumbert (1980) in her doctoral thesis *The Phonetics and Phonology of English Intonation*. Her treatment of intonation has been widely referred to as the basis of intonation research nowadays and it was named by Robert Ladd (1996) as Autosegmental-Metrical model (AM model). Before the review of this model, a brief introduction of Liberman and Prince's (1977) work on English stress will be presented for a better understanding of the AM model.

In light of Chomskyan revolution against structuralist approaches to morphology and syntax, Liberman and Prince (1977) introduced a generative approach to stress (called metrical phonology), by which stress is considered as a reflection of a hierarchical rhythmic structure that governs the syllables, words, and syntactic phrases within a sentence, instead of treating it as an inherent property of segments or syllables. They propose a formal system in which stress is defined by metrical tree structures whose nodes branch binarily into s (strong) and w (weak). Since the assignment of s and w nodes is governed by two rules applied to word/compounds level (Lexical Category Prominence Rule) and sentence level (Nuclear Stress Rule), the prosodic stress is thus predictable at the surface structure. The core essence of this theory is that stress levels are relative rather than absolute.

Following this philosophy, Pierrehumbert (1980) analysed intonation as a sequence of pitch targets that are made up of two abstract tones: high (H) and low (L) tones, either on their own or a combination of the two. Similar to s and w representing relative stress levels in metrical phonology, H and L are also relative tones in the AM theory. The significant pitch targets in the AM model include *pitch accent*, *phrase accent*, and *boundary tone*. The pitch accents are localised on the metrically generated stressed syllables, and this is the metrical aspect of the AM theory. The autosegmental aspect is responsible for the association of H and L tones to

the metrical structure. Pitch accents can be mono-tonal and bitonal, and the tone target associated with the stressed syllable is added with an asterisk (*), resulting in H*, L*, L*+H, L+H*, H*+L, and H+L* (Pierrehumbert and Hirschberg, 1990; Ladd, 1996). Phrase accent and boundary tone are both presented as either H or L. The difference between the two is the phrase accent occurs at the end of the intermediate phrase while the boundary tone would appears at the end of the intonational phrase (the tone unit/group in the British school) (Pierrehumbert and Beckman, 1988). The transcription of phrase accent is also slightly different to that of boundary tone in terms of the diacritics attached to H and L; the former is H- / L-, while the latter is H% / L%. Each Intonational phrase should have at least one pitch accent. When there is only one pitch accent, it is called the nuclear pitch accent (the nucleus in the British school). When there are more than one pitch accent in an intonational phrase, the last pitch accent is the nuclear pitch accent, while any before is the pre-nuclear pitch accent (compared to the head in the British school) (Ladd, 1996). This kind of analysis aims to create a system of underlying phonological representation that is able to generate all the possibilities of the surface phonetic realisations of the English intonation, and now its influence has spread over the research of many other languages.

2.1.3 Comparison between the British and American approach

As reviewed in the previous two sections, the fundamental difference between the British and American school of intonation research can be identified as the way they conceive of the primitives of intonation. The British school treats intonation as constructional bits: pre-head, head, nucleus, and tail, each of which has its own contour that is independent of other bits. So pitch contour such as fall, rise, and fall-rise cannot be decomposed into smaller units. The nucleus (or tonicity) is the most important part of intonation in the British school and it is the only connection between tones and the text (Gut *et al.*, 2007), whereas the American school recognises pitch levels (H and L) as the constituent of the significant prosodic events (tone

movement). In the AM model, the nuclear pitch accent is no more important than the pre-nuclear pitch accent; it is simply the last pitch accent within an intonation phrase.[3]

Another difference between the two schools can be traced back to the original aims of their analysis of intonation. The British school was and has been pedagogy-oriented, and later scholars are always building upon the previous achievements (e.g. Palmer, 1922; Kingdon, 1958; O'Connor and Arnold, 1961, Brazil, et al., 1980; Cruttenden, 1990; Wells, 2006). The American school, or more precisely the AM model,[4] drawn on generative linguistics, intends to create a system that could transcribe the phonological inventory of intonation constrained by a particular language's phonological rules (Pierrehumbert, 2017). The annotation system (ToBI) derived from the AM model (Silverman *et al.*, 1992; Beckman and Ayers, 1997) has been extensively used not just in English, but for many other languages too and modified into their own fashion, e.g. C-ToBI for Mandarin Chinese. This annotation system renders it friendly to the instrumental analysis and makes the cross-linguistic comparisons easier. Applied linguists and language teachers, however, are not likely to be attracted by AM theory; they tend to sustain the use of the British nuclear-tone theory which bears a longer history in English teaching (Levis and Wichmann, 2015), presumably because "these theoretical arguments are far beyond the needs of most teachers of English. They are also largely unrelated to what happens in the classroom" (Levis, 2005: 343).

AM theory has long been criticised for its neglect of intonation meaning of those generated typical contours in English (Cruttenden, 1997). Although Pierrehumbert and Hirschberg (1990) had made an effort to explain the linguistic meanings conveyed by those intonation contour patterns, this problem seems to remain thorny. The British school, on the other hand, has been questioned as to how many pitch contours are significant in English, how to map pitch contours onto the words and syllables, etc. (Chun, 2002).

[3] In Pike's (1945) system, however, nucleus was also more important than other parts of pitch movements.
[4] Earlier American models, such as Pike's (1945) was actually pedagogy-oriented and has been used in some recent textbooks, e.g. Celce-Murcia *et al.*, 2010, but later analyses were more approached to theoretical rather than applied linguistics.

Despite the differences in the foregoing discussion, the nuclear model and the AM model inevitably shared something, which in Levis's (2005) and Cruttenden's (1997) opinion, might have outweighed their dissimilarities. First, they both agree that it is the overall tune of the entire phrase being of importance. Second, both insist that intonational meanings are generated from an interaction of phrasing, pitch accent placement and tone assignment. Third, both distinguish far more meaningful contrastive tune patterns than what have been taught in the textbooks. Apart from these, the AM model has in fact been influenced by the British school by means of incorporating the notion of pre-nuclear vs. nuclear pitch accent (Ladd, 1996, 2008), and the transcription of a particular pitch pattern by ToBI system can find its counterpart in the British tradition, for instance, the nuclear falling tone at the sentence final position can be represented as H*L-L%.

2.1.4 Meanings and functions of English intonation

Wells (2006) proposed six functions of English intonation that are achieved by phrasing (tonality), nuclear pitch accent location (tonicity) and choice of tones, most of which are consistently researched by other scholars too, i.e. attitudinal function, grammatical function, focusing function and discourse function. This section will be focusing on these four functions with reference to other research when needed, while leave the psychological and indexical function untouched because they are beyond the realm of the current study.

The first and probably the most intuitively recognised function of intonation is *attitudinal* (or *affective*) function that had been comprehensively described by Pike (1945) and O'Connor and Arnold (1961). According to Pike, intonational meanings were superimposed on lexical meanings as such that a speaker can manipulate the intonation of the utterance to express his/her emotions and attitudes that are beyond the words, and that are expected for the hearer to perceive. Similarly, O'Connor and Arnold constructed ten pitch contours that were thought to be of contrastive attitudes, assuming that intonation functioned mainly attitudinally. Their work has been found convenient for teaching language learners (Zhang, 2003), but the

overemphasis of intonation's attitudinal function might lead to a mechanical use of intonation patterns in some cases that only a change of words could result in a different affective meaning. Brown et al. (1980, 2015) also suggest that voice quality might be a better indicator of attitudes than intonation; for example, final rises, together with a "kindly" voice quality, are sort of associated with kindly and encouraging attitude among the interlocutors, but if someone changes the voice quality while using a rise tone, the meaning might change to "politeness" and "hectoring" (Brown, 2015: 22). Attitudinal function is also concerned with the extent to which intonational expression of attitudes is linguistic. A person using an "angry" intonation may be truly experiencing that emotion, or consciously use this particular intonation pattern to make the listener believe that s/he is angry irrespective of his/her real inner state. With its psychological involvement, attitude is deemed to be of great difficulty in mapping to the linguistic use of intonation (Nolan, 2006: 446).

The second function of intonation that has also been widely commented on is *grammatical* function. What intonation does to the grammar in speech might be analogous to what punctuation does in writing. Wells (2006) categorizes two types of grammatical functions that learners of English can grasp by manipulating the intonation. 1. Grammatical units such as clause and sentence can be demarcated by tonality (phrasing); 2. sentence types (question/statement) and ambiguous syntactic structures can be distinguished by tone (pitch accent). Chun (2002) also favours this dichotomy of grammatical functions that intonation encodes. The first one seems to be less challenged, as in an examination of the degree to which intonation phrasing co-occurs with syntactic structures, Cruttenden (1997) has found that the clause is the major unit that matches the intonation phrase among many other smaller syntactic units, e.g. adverbials, noun-phrase subjects, vocatives, topicalized subjects, etc. This is specially legitimate in read speech (Timkova, 2001; Ramírez Verdugo, 2003), but in spontaneous speech, the relatedness between intonation boundary and grammatical boundary might be less obvious (Levis and Wichmann, 2015: 151). For the second type of grammatical function, there seems to be a tendency in English towards the falling vs. rising distinction (or

L% vs. H% boundary tone) in relation to statements and questions respectively (Pierrehumbert and Hirschberg, 1990; Liu *et al.*, 2013). Although it is not always the case, L2 learners of English should at least know that if they are not seeking answers from their listeners, a falling tone should be better than a rising tone to use, and if they are looking for more responses, a rising tone is ideal, as in the case of tag question, e.g. It is raining, isn't it? As for the syntactically ambiguous structures, intonation plays a major part in the delivery of a correct meaning that anyone wants his/her listener to perceive, as exemplified in the following relative clauses (adapted from Nolan, 2006: 441). The pre-nuclear accent is underlined; the nucleus is in bold. Intonational phrase boundaries are marked with |.

(1) The Norwegians who are rich enjoy life to the full.

(a): The Norwegians who are **rich** | enjoy life to the **full**.

(b): The Nor**we**gians | who are **rich** | enjoy life to the **full**.

If the message someone wants to convey through above sentence is (a) "only rich Norwegians enjoy life", the restrictive relative clause will be adopted as the phrase boundary would be put in between "rich" and "enjoy". If it is said with a non-restrictive relative clause (b), an extra boundary will be inserted after "Norwegians", meaning that all Norwegians are rich, and they enjoy life. The non-restrictive phrase "who are rich" acts like a parenthesis in (b) that would normally receive a compressed pitch range (Brazil, 1997). It is noted that along with the insertion of one more intonation phrase, the nuclear pitch position and the nuclear tone choice are in accordance changed. In sum, grammatical structure can be used as a prediction for intonation, whereby learners can determine whether or not they cut their speech at particular syntactic boundary, whereas structures like relative clauses need to be explicitly taught (Nolan, 2006).

The third function is *focusing*, also known as *informational* or *accentual* function. Brown et al. (1980) pointed out that informational function is "one of the currently most discussed functions of intonation" (p.27), and Wells (2006) acknowledges it as "one of the most

important…and perhaps…most readily taught in the EFL classroom" (p.11). This function is implemented by tonicity (placing of nuclear pitch accent) and by the placement of other pitch accents, alongside the selection of a type of nuclear tones. Speakers always chunk their utterances into smaller pieces each of which entails only one piece of information by making one word or two stand out as the most salient part of the chunk, the "information focus" or "foci" (Halliday, 1967b). From the perspective of communication, nucleus will hence be the most important part within an intonation unit as it signals the information centre (Bolinger, 1968) that the speaker employs most of his or her effort to produce and that the listener attends to. Normally, English words that carry new or unpredictable information in an utterance gain more prominence than the given/old or contextually deduced information (Halliday, 1967b; Chafe, 1976). The following example remarks on the use of intonation to decode different informational meanings (adapted from Nolan, 2006: 442).

(2) I'll be there about five.

(a): | I'll <u>be</u> there about **five**. |

(b): | I'll <u>be</u> there **about** five. |

(c): | **I'll** be there about five. |

If this utterance is said with a neutral meaning, i.e. stating a fact, the nucleus will fall on "five", the last content word as shown in (a). If it is "above", instead of "five", that is emphasised by the speaker, it indicates that five o'clock is a given/old information that has been informed or implied in previous contents, and the speaker is trying to contrast "about" to maybe "before" or "after", suggesting that there might be a misunderstanding between the interlocutors and the speaker attempts to correct it so that the hearer can update the information that he or she is expected to perceive. The same applies to case (c) where "I'll" is a contrastive information to "He'll", "She'll", etc. One might confuse the conception of "new" with "contrastive" as contrastive information could be to some extent a type of new information. In fact, researchers can hardly reach to a consensus of the definition of the two,

e.g. Halliday (1967a: 211) conceiving new information as either "cumulative to or contrastive with what has preceded", whilst Chafe (1976) argues that contrastive information is a category independent of new information. There is no denying that highlighting particular parts of the speech is very important for successful and smooth communication. As long as the information that a speaker intends to convey is uncontroversially delivered by means of proper placement of pitch accents and tones, whether the information is new, or contrastive, or both should not bother the speaker.

In regard to the phonetic realisation of (nuclear) pitch accent, there is less agreement. Basically, the distinctive correlations that most researchers have consistently reported include higher amplitude, longer duration, more expanded F0 range, and more precise articulation of vowels than that of unaccented syllables (Cooper *et al.*, 1985; Sluijter and van Heuven, 1996; Cruttenden, 1997; Gut *et al.*, 2013). Apart from those on-focus adjustment of phonetic details, Xu and his colleagues have also noticed the changes of post-focus words whose F0 and intensity are largely reduced, while pre-focus words are rarely affected (Xu and Xu, 2005). This phenomenon is called post-focus compression (PFC) (Xu *et al.*, 2012) and has been evidently present not just in English, but Mandarin Chinese (Xu, 1999; Wang and Xu, 2011), Korean (Lee and Xu, 2010) and many other languages. Perception studies have found that native English speakers might adopt all on-focus cues (Wennerstorm, 1994) as well as PFC (Prom-on *et al.*, 2009) to receive information signals. All these suggest that L2 learners who either fail to correctly place the pitch accents, or fail to produce proper phonetic qualities of pitch accents and their surrounding syllables, or even fail to do both are very likely to encounter a communicative breakdown.

Apart from information structure that is intimately bound with intonation, pragmatic meaning, realised mainly by tone is considered as the other aspect subsumed under the informational function by Wells (2006). The distinction of fall and non-fall (rise and fall-rise) is at the centre of discussion. The common proposition is fall signals finality (Wichmann, 2000) or closure (Cruttenden, 1986, 1997), while non-fall refers to nonfinality or openness. Wells

(2006: 75) uses listing to demonstrate the different use between falls and non-falls, e.g. You can have **cof**fee ｜ or **tea**. In this case, the nuclear falling tone on "coffee" in the first intonation unit might imply that there is no more choice than coffee so that the listener would not expect more options to come. But apparently "tea" is an alternative in this case. As a speaker, a rising tone would be the optimal choice for "coffee", and the same applied to "tea". If a nuclear rising tone is assigned to "tea", it conveys a message that "coffee" and "tea" are just two options among many other things yet unmentioned. Fall-rises, in Wells' (2006) definition, have an additional meaning on openness, which is called implicational fall-rise, often equivalent to "but". This is an example from Wells (2006: 27):

(3) Who's that?

Well I know her **face**.

A fall-rise on "face" might imply that "but I can't remember her name". Of course, the speaker can make this meaning explicit by saying it out, but whether it is implicit or explicit, the implicational meaning is embedded in the fall-rise tone. The fall/non-fall distinction has been further explained by Gussenhoven's (2002, 2004) production code under which high endings (non-falls) marks continuation;[5] low endings (falls) express finality and end of turn. If L2 learners who are unable to produce or comprehend proper nuclear tones that distinguish the pragmatic meanings might experience misunderstanding or confusion in interactions.

The last function covering beyond the sentence level is discourse function. By manipulating prosodic features, speakers can maintain a cohesive narrative and manage turn-taking with their interlocutors. The most commonly exploited prosodic features for discourse meaning are gradient in nature, such as pitch range, tempo, and loudness (Levis and Wichmann, 2015). Enlarged pitch range with an extra high pitch on the first accented syllable indicates a large part a change of topic (Wichmann, 2000). A pause is often present too before the start of the

[5] One of Gussenhoven's (2002, 2004) three biological codes guiding phonetic implementation of intonation that conveys paralinguistic meanings. Production code refers to the energy pushed out when speaking that coincides with exhalations phases of the breathing process. The other two codes are frequency code (variations in individual's larynx corresponding to variations of F0) and effort code (energy required for speaking).

new topic in read speech, but in extemporaneous speech, the pause could be replaced by a sudden acceleration of speech, which is named “rush-through” (Couper-Kuhlen and Ford, 2004; Local and Walker, 2004). On the contrary, a compressed pitch range at the beginning of an utterance signals continuation of the topic or a tight connection to previous utterances. A typical example of this kind of usage of intonation is in parentheses. A parenthesis tends to be articulated with lower pitch range and faster speed compared to its surrounding utterances, but in some rare cases, the pitch range might be expanded (Dehé and Kavalova, 2007).

In summary, intonation functions reviewed so far are not mutually exclusive of each other, but interrelated and overlapped, as it seems difficult to find a one-to-one match between intonation form and function. For example, the rise/fall distinction can ascribe to grammatical function, pragmatic function, or discourse function. The high/low pitch can either deliver attitudinal meaning or discoursal meaning. Overall, it is not the function to which intonation links matter; it is how semantic and pragmatic meanings are conveyed and unambiguously understood between speakers. L2 learners are faced with a number of options for intonation forms. The best method to teach intonation is perhaps not overemphasising the correlation between one particular form and one particular function. L2 learners should be encouraged to use whichever intonation they want to use, as long as they use it in a way that precisely reflect what they are thinking and what message they actually want their interlocutors to perceive.

2.2 Mandarin Chinese intonation

This thesis targets at Mandarin Chinese as the L1 because of the relatively extensive research on it compared to other Chinese dialects. Chinese is a large language family under which seven major varieties (dialects) are recognised. These dialects are mutually unintelligible but share the same written language and grammatical rules (Lin, 2007). In the early twentieth century, a language reform campaign was led by a panel of renowned linguists and scholars who endeavoured to standardize the northern dialect into a national language, attempting to

ease the difficulty in communication across regions (Duanmu, 2002). This standardization has made Mandarin Chinese (also called Standard Chinese, or Putonghua) the official language in China for almost a century and has become compulsory in schooling. The intonation system introduced in this section is strictly confined to Mandarin Chinese (MC), and will be preceded by a brief description of the stress and tonal system of MC, as these three aspects of prosody are closely intercorrelated and manifested by similar acoustic dimensions.

2.2.1 Tone and stress of Mandarin Chinese

Chinese is a tonal language (Chen, 1999; Duanmu, 1999; Yip, 2002). In addition to consonants and vowels, tone in Chinese also serves to differentiate lexical meanings. Phonologically, the syllable is the tone bearing unit in Chinese—identical syllables vary semantically with different tones (Wang, 1967; Chao, 1968). Tone is manifested by the fundamental frequency (F0) (Ladefoged, 2001). Tone in MC has two types in terms of pitch movement. One is the level tone, meaning that the pitch remains stable throughout the syllable and there is only one level tone in MC (Tone 1). The other three lexical tones are contour tones which means that the pitch changes from the start to the end of the syllable (Tone 2, 3, 4) (Lin, 2007, p. 92). Tone 1 is a high level tone that is normally transcribed as 55 in Chao's (1930) digit or HH.[6] Tone 2 is a mid-rising tone (35/MH), Tone 4 a high-falling (51/HL). Tone 3 is slightly complicated. It starts low and goes to an even lower dip before rising to the mid-high level (214/LH), but when it is followed by other tones, it keeps at the lower level without the rising bit, a phonological phenomenon called tone sandhi (Zhang, 2007),[7] resulting in the transcription as 21 or LL.

[6] Chao's (1930) digit is the numeric transcription of the pitch values of Mandarin Chinese tone system with 5 representing the highest pitch and 1 the lowest.
[7] Tone sandhi is a phonological phenomenon whereby an underlying tone might be changed to another by the influence of its adjacent tones (see Lin 2007 and Duanmu 2003 for detail description of tone sandhi of Mandarin Chinese).

Apart from the four phonemic tones, there is a fifth tone (T0) in MC whose occurrence is highly contextualised. When any phonemic tone loses its original tone, it becomes T0, the neutral tone. It has to be preceded by at least one syllable which carries a phonemic tone. The phonetic manifestation of the neutral tone is shorter in length and weaker in loudness compared to its underlying tone with reduced vowels. Some argued that T0 is toneless (Lin, 2007; Duanmu, 2014), while others object and assert that the pitch level of T0 depends on the preceding tones (Cheng, 1973; Yip, 2002). Despite the mixed conceptions of the F0 dimension of the neutral tone, its other phonetic manifestations seem to mirror that of unstressed syllables in English, a fact stirring up a discussion of whether or not MC has lexical stress. A strong piece of evidence provided by the proponents for the existence of word stress is that MC tends to be minimally two syllables long, a foot-like structure which is left-headed just as in English (Duanmu 2002: 140). To make a monosyllabic morpheme into a disyllabic word, either the monosyllabic morpheme is reduplicated, like di (T4) 'younger brother' and didi (T4T0) 'younger brother', or another morpheme is added with the same meaning or is meaning-free, such as xue (T2) 'study' followed by xi (T2) 'study' to make the word xuexi 'study' (T2T0) (Lin, 2007: 224-225). In both cases when the second syllable is neutralized, the perception of stress and unstressed syllables seems to be no challenge for native speakers (Chao, 1968; Wang and Feng, 2006), whereas in situations when no neutral tone is present, the perception of the stress and unstressed distinction tends to be difficult (Chen, 2000). Duanmu (2002) contends that the difficulty of perceiving stress in Mandarin is due to the limited manipulation of F0—an important acoustic cue for stress—which has been adopted for distinguishing lexical tones. However, a more recent study has found that Mandarin stress is associated with high pitch in tones, indicating that F0 might play a role in stress perception by native speakers (Wang and Chu, 2008).

2.2.2 Intonation of Mandarin Chinese

Unlike English in which F0 can be freely exploited for the encoding of intonation, F0 in MC has its dominant role in defining lexical meanings already; thus it cannot be as freely manipulated as in English for intonational use, making the research of Chinese intonation particularly difficult because of its inherent complexity (Cao, 2000). In fact the debate on how to separate intonation from tone in Chinese is still ongoing, but undisputed is Chinese does have intonation (Chao, 1933; Shen, 1989; Xu and Wang, 1997; Cao, 2004; Lin, 2007; Lin, 2008), although in many cases, a syntactic rather than prosodic strategy will be used to deliver contextual meanings, so as to "avoid the potential conflict between tone and intonation…like many other tone languages (especially Asian tone languages)" (Lin, 2007: 228). For instance, MC employs sentence-final particles to denote sentence types as in "Is she a lecturer?" (ta shi jiangshi *ma*); "ma" is the question marker. Without "ma", the sentence becomes a statement, leaving other morphemes intact. Another example is "ba" which indicates the meaning of "supposition" or "solicit agreement" (Lin, 2007: 229) as in "Come with me, please!" (gen wo qu *ba*). With the use of these particles, F0 change is likely to be restrained to the minimum for intonation use.

Traditionally, Chinese linguists have believed that there are underlying intonation patterns covering the whole sentence that is superimposed on the tonal patterns of each syllable, given the circumstance that lexical tones cannot lose their pitch pattern to an unrecognizable state (e.g. the rising Tone 2 cannot change to the falling T4). Shen (1989) proposed three intonation patterns designated to statements (Type I), questions (Type II), and questions beginning with question words (Type III). In general, question contours (Type II and III) had higher initial pitch register than statements. The difference between the Type II and Type III question fell on the sentence-final pitch register, the former ending high and the latter low, concluding that it was the pitch register instead of the movement that mattered in Chinese intonation. Some other studies proposed a pattern of declination over the entire utterance of statements—the high level tone at the beginning of the utterance was the highest, while later high tones were

gradually lowered, and the high tone at sentence-final position was the lowest (Shih, 2001; Wang, 2003). However, Xu (1999, 2001) had found an interdependent effect of low tone and focus/new topic on the declination phenomenon, therefore rejecting the proposal for the explicit global intonation contour for MC.

Under the overwhelming influence of the AM theory, a growing body of research has started to examine Chinese intonation in the same framework. Controversies never cease to appear in regard to the type and inventory of phonological components of MC. The boundary tone seems to be at the centre of debates. Researchers who are against boundary tones claim that the final lexical tone retains its shape regardless of sentence type, but either compress or enlarge the pitch range of that particular tone for declarative and interrogative sentences respectively. Also, the overall pitch register of the entire utterance tends to be higher for questions than for statements (Ho, 1977; Shen, 1989; Kratochvil, 1998). On the contrary, Lin (2004a, 2006, 2008) argues that the intonational distinction between questions and statements in Mandarin is crucially realised by an H and L boundary tone added onto the final one or two syllables of the utterance without the change of their basic lexical tone. It means that for interrogative utterances, the high boundary tone applies to the sentential terminal by raising the ending pitch point and/or the starting-point of the syllable (with smaller degree) and making the tone contour steeper. The low boundary tone makes the starting pitch point lower, resulting in a milder slope of pitch contour of the final syllable. Some other researchers also agree with an inclusion of boundary tone in Chinese intonational phonology, but its interpretation varies according to the linguistic domains and methodological treatments that they controlled for (e.g. Schack, 2000; Lin and Li, 2011). Nonetheless the notion of boundary tone in Chinese analyses seems to deviate from the AM theory in which the boundary tone indexes demarcation of intonation phrases, whereas in Chinese intonation, boundary tone for the majority of the cases solely associates with sentence types (Jiang and Chen, 2011). Regarding the phonetic application of phrase boundary in MC, three markers have been consistently found: pause, boundary syllable lengthening, and boundary syllable

strengthening (Xu and Wang, 2009; Cao, 2012). The first two seem to echo the boundary indicators of English, although the syllable lengthening in MC includes the syllable at the onset (post-boundary) and offset (pre-boundary) of the phrase, and a difference has been discovered between the two positions (Cao, 2005). Final syllable lengthening normally happens on the rime of that syllable, while post-boundary lengthening occurs on the consonant (Cao, 2005). Cao and her colleague (2006) have also found that the articulation of the initial consonant at post-boundary and the rime at pre-boundary tend to be strengthened by more complete and robust articulation.

More agreement exists for the pitch accent in the analysis of MC. As mentioned in 2.2.1, phonologically MC has left-headed foot structure, but its phonetic manifestation has been inconsistently reported due to the complex interaction of stress and tone. As a result, linguists working on Chinese intonation tend to direct their attention to focusing or information structure, a principal functional use of pitch accent, to explore the interpretation of pitch accent in MC. In line with English, information centre (newness, contrastive, and emphatic focus) becomes phonetically salient in MC when syntactic cues are absent (Xu, 2004b; Li, 2012). What makes MC accents different from English is the way they are modulated. Accumulated evidence is elongated syllable duration, increased pitch range and heightened pitch register of the focused word (e.g. Xu, 1999; Jia *et al.*, 2008; Lin and Li, 2011; Wang *et al.*, 2018), while the mechanism behind those surface realisations seems to be diverse when different locations and domains of the focus are sought. For example, in Lin and Li's (2011), the expanded pitch range of the focused syllables were achieved from the raised high pitch of Tone 1, 2 and 4 and lowered low pitch of Tone 3, whereas in Xu (1999), the lowered low pitch also applied to Tone 2 and 4. In addition, Xu (1999) found that the scale of the pitch range expansion was affected by the position of the focus: sentence-initial and medial focus was seen with larger expansion than those in sentence-final position. Jia et al. (2008) extended Xu's study to longer domains of Chinese utterances and the focused words (disyllabic long) included all 16 possible combinations of lexical tones (e.g. T1T1, T1T2, T1T3, T1T4, T2T1,

T2T2…T4T4). They found that irrespective of the focus position, the pitch expansion, however, was a result of heightened high pitch of all four phonemic tones, and the high pitch raising applied to both syllables in the focused words. The accompanying perception experiment in the same study revealed that the more high pitches a syllable has within the focus domain, the higher chance that the syllable will be perceived as being stressed. This provides an empirical basis for the existence of metrical structure of MC, and it has been confirmed by a study in which sentences with multiple focuses (Jia *et al.*, 2010) were targeted to investigate the phonetic correlates of pitch accents and the distribution of prominence and its relation to focus. Based on a robust statistical analysis, the authors argued that the distribution of focus and prominence was likely to be asymmetric. When the sentence has dual focus, prominence co-occurred on both focused words, but the right one stood out more prominently than the left one in the form of wider F0 raising. Interestingly, sentences with multiple focuses would only exhibit one prominence at the rightmost focused position. Drawing on such results, they proposed that MC might have the pre-nuclear and nuclear distinction as English does in which nuclear accent is obligatory and exhibits most dramatic pitch changes while pre-nuclear accent is optional and of less F0 turbulence.

Another significant finding related to focus in MC is the post-focus compression (PFC). F0 height and range is highly likely to be lowered and compressed on the syllables immediately following the focus (Xu, 1999; Liu and Xu, 2005; Liu and Xu, 2007; Jia *et al.*, 2008; Wang and Xu, 2011; Xu *et al.*, 2012; Xu, 2017; Wang *et al.*, 2018). The extent to which the PFC applies on the utterance has unfortunately rarely been touched, as most of these studies are restricted to short utterances whose post-focus domains covered only one or two syllables. A handful of studies such as Jia et al. (2008) offer evidence by analysing longer sentences and assert that the PFC compresses the pitch range of every syllable in post-focus position when the focus is localised on the beginning and middle of the utterance. Further investigation by Wang et al. (2018) has found that PFC modulates the pitch range across phrase boundaries; in other words, the prosodic boundary seems to play no role in eliminating the effect of the PFC

on the syllables beyond the scope of current phrase where the focus is. The PFC has been seen in English too, but the difference lying in between the Chinese and English has been identified. In English, the pitch range and height of post-focus syllables in statements were compressed and lowered, while those in questions were compressed but raised (Liu and Xu, 2007), indicating that information structure (or focus) interacts with sentence modality that is reflected in intonational encodings of meanings. This interaction has been in fact found for Mandarin Chinese too but in a way that only pitch range is concerned. Simply put, the pitch range expansion on the focused word is larger for questions than for statements. The PFC applies to both statements and questions but to a lesser extent in questions than in statements (Liu and Xu, 2005). Another difference of PFC between Chinese and English worth mentioning is that in English questions, the F0 in post-focus domain gets higher all the way to the end of the utterance, and the F0 maximum is even higher than that in the pre-focus position. But in MC, with the effect of PFC, the pitch height after the focused word would by no means get higher than that in the pre-focus words (Xu, 2017).

Besides pitch-related variations, durational effect is another reliable measure in distinguishing accented syllables from the non-accented in MC and some researchers have even argued that duration is more faithful than pitch to signal accentuation in MC when different types and degrees of focuses are concerned. Greif (2010) have found that the phonetic commonplace in the semantic correction and pragmatic correction in MC is the lengthened duration of the focused words compared to non-focused words, and the difference between the two is the pragmatic correction increases the F0 span additionally while the semantic correction does not. Chen's (2008b) finding seems to be partially supportive of Greif's (2010), though the target in her analysis was different scales of corrective focuses. She found that the moderately corrective focus was longer in duration and wider in pitch range than non-focused words, whereas the full corrective focus only exhibited longer duration than the moderately corrective focus, leaving the pitch range intact. A more recent study (Bi *et al.*, 2016) has shown similar results when the position and types of focus interact. They discovered that in

the subject position, words for corrective focus tend to have longer duration, wider F0 range and larger intensity than givenness, while no significant difference is found between informational focus and givenness. Between corrective focus and informational focus, only durational difference is found. In the object position, F0 difference is not found between any types of focus, indicating unreliable F0 range as an indicator of information structure.

To summarise, Mandarin Chinese and English exhibit great differences in terms of the phonetic realisations of intonation while the intonational functions large overlap such that both can signal sentence modality, disambiguate syntactically confusing meanings, highlight information centre, and express attitudes and emotions. Although research on Chinese intonation has seen its own route from the beginning and is on its way to gradually converging to the mainstream intonation research by an increasing number of dedicated researchers, the comparisons between English and Chinese intonation are still difficult under the same framework of intonational phonology plausibly due to their inherent diversity in prosodic typology and the relatively limited research on Chinese intonation which is addressed within the same theoretical framework as most of the work on English intonation. Section 2.1 and 2.2 have made an attempt to review the literature on both English and Chinese intonation that seem to be comparable and that might provide insights in the explanation of intonational performance by native English and Chinese learners in later chapters.

2.3 L2 acquisition of intonation

Research on the L2 acquisition of intonation has long been marginalised within the field of L2 speech acquisition, although its resurgence has triggered an accumulation of literature on the investigation of perception and production of L2 intonation (mainly confined to English) from the phonetic and phonological perspectives by L2 learners with multiple L1 backgrounds. This has also been reflected by the fact that the most influential theories of L2 speech acquisition such as the Speech Learning Model (SLM) (Flege, 1995; Flege *et al.*, 2003) and

the Perceptual Assimilation Model (PAM/PAM-L2) (Best, 1995; Best and Tyler, 2007) have only included segmental aspects to explain how and how well non-native sounds are learnt. While the need for modelling L2 suprasegmental acquisition has been recognised (see Gili Fivela, 2012; Mennen and de Leeuw, 2014) and efforts have been made by So and Best (2008, 2010, 2011, 2014) to expand PAM-L2 to account for L2 acquisition of lexical tones (PAM-S), it is only until very recently, that a working model of L2 intonation acquisition has been proposed, by Mennen (2015)—the L2 Intonation Learning theory (LILt)—drawn on her earlier works (e.g. 1999, 2004, 2007). This section will be focused on detailing the LILt model with reference to the existing empirical studies on L2 learners' performance of non-native intonation, particularly targeting Chinese learners of English. Other models (i.e. SLM and PAM/PAM-L2) will be referred to when they are relevant to explaining the underlying difficulties of acquiring L2 intonation, but the details of these models will be omitted as they are not of direct interest to the current study.

2.3.1 L2 Intonation Learning theory (LILt)

The basis of this model is to account for the difficulties that L2 learners might encounter in the production of non-native intonation along four dimensions, the systemic or phonological dimension, the realisational or phonetic dimension, the semantic dimension, and the frequency dimension. According to Mennen (2015), the systemic dimension refers to the inventory and distribution of categorical phonological elements (e.g. pitch accents, boundary tones, etc.). The realisational dimension is the phonetic implementation of these categorical systemic elements. Unlike segments, intonation forms are inevitably associated with functions. Thus the third dimension of LILt—the semantic dimension—concerns the semantic and pragmatic meanings delivered by these phonological categories, whereas the frequency dimension deals with how often a particular intonation element is used in a language. Taking these four dimensions into account, it is possible to predict the difficulties that might arise for

learners with various L1 backgrounds by systematically comparing the similarities and dissimilarities between their L1 and the target L2 on these dimensions.

The LILt makes four important assumptions that seem to correspondingly fit into the SLM (Flege, 1995) and PAM-L2 (Best and Tyler, 2007). The first and foremost is that learners' difficulty in the production of L2 intonation is partly attributed to their perceptual bias of intonational categories. In both SLM and PAM-L2, L2 segments can either be perceptually assimilated to L1 categories where production deviances can be found among L2 learners, or be dissimilated away from the L1 categories where new categories will be formed and little difficulty should be posed for L2 learners. Whether an assimilation or dissimilation process is triggered depends primarily on the degree to which the L2 segments resemble the L1 category phonetically. However, as conceded by Mennen (2015), it is far from clear to determine the categories of intonation due to the intrinsic complexity of the relationship between intonational form and function. Therefore, it is suggested that information or meaning references are better accessible to listeners when making judgments of their perception of intonation categories (Gili Fivela, 2012).

The LILt also acknowledges the importance of age-related factors (i.e. age of arrival, age of leaning) contributing in L2 learners' proficiency of intonation, but whether the influence of age is parallel at all four dimensions needs to be empirically investigated (Mennen, 2015).

The third important assumption from LILt is, concurring with SLM and PAM-L2, L2 learners' production of intonation gradually approaches to the L1 norms as their L2 proficiency in general increases at least in certain dimensions (see Mennen *et al.*, 2010). Some learners can even reach perfect production like L1 speakers (Mennen, 2015). But there is doubt that all four dimensions are open to successful acquisition.

Both SLM and PAM-L2 posit that L1 and L2 segmental categories interact bi-directionally in a common phonological space (Flege, 1995; Mennen, 2004) where the result of interaction is either merging or polarizing the phonetic features of L1 and L2. If merging occurs, learners are likely to produce the features with values in between L1 and L2, whereas if polarizing

happens, learners tend to overshoot the values so as to maximise the difference of phonetic features between the L1 and L2. The LILt claims that L2 learners might experience merging or polarization in the process of learning intonation too (see, Mennen *et al.*, 2014), although some might be capable of differentiating the two systems and achieve success ultimately (Mennen, 2015).

2.3.2 L2 learners' production of intonation

It is gratifying that the research on L2 learners' performance of foreign intonation is being substantiated in recent decades, although neither the size nor the scale of it is comparable to that of L2 segmental studies. The findings of these studies beyond all doubt have offered an opportunity to discover the difficulties and the potential causes of these difficulties behind the scenes, to verify the proposed acquisition theories, and to enlighten the teaching and learning of L2 intonation. This section is an overview of the previous empirical studies which are focused on the production of L2 intonation.

There is abundance of evidence showing that L2 learners have problems in producing native-like intonation phonetically and phonologically (e.g. Mimutsu, 2000; Ramírez Verdugo, 2003; Toivanen, 2003; Mennen, 2006; Chen, 2008a; Ding *et al.*, 2012b; Gut *et al.*, 2013; Li and Post, 2014). Learners of different L1 backgrounds might show different problematic production patterns. Even though some patterns universally appear among different groups of learners, the level of difficulty with which they acquire particular intonational features might vary. All these findings are thought to be more or less due to the influence of their mother tongue.

Many Chinese learners of English, at the centre of the discussion, have also been found with divergent manipulation of phonetic cues in the assignment of pitch accents and phrase boundaries from native norms. For example, Ji *et al.* (2009) revealed that compared to native speakers of American English, MC learners rarely used a low-rising tone on nuclear words on

the medial part of the sentence in yes-no questions. They favoured more of a high level tone or falling tone on a medially positioned nuclear word and a falling tone on the sentence-final nuclei. As for native speakers, a low-rising tone seemed to be optimal for the nuclear word in yes-no questions regardless of the position of the nuclear word. This finding was supported by Wang *et al.* (2011a) and Shao *et al.* (2011) who found that Chinese learners with different dialects were in general reluctant to use lowered F0 to signal pitch accents, although F0 dimension was very likely to be their top phonetic manipulation of accentuation. They tended to rely on raised and/or expanded F0 to mark this type of information, apart from which lowering F0 was another typical choice for native speakers as found in Ji *et al.* (2009). However, the manipulation of duration and intensity of accented syllables by L1-Mandarin Chinese were similar to native speakers', as both groups lengthened the vowels and increased the intensity to make the accents prominent. These consistent findings regarding to the phonetic realisation of pitch accents were also uncovered for Chinese learners of other West Germanic languages, such as Dutch (He *et al.*, 2011) and German (Ding *et al.*, 2012b).

As reviewed in section 2.2.2, the way that Mandarin Chinese realises narrow focus closely resembles English by lengthening the duration and heightening the intensity of the focused syllable. Due to the limitation of the change of the lexical tone, the manipulation of F0 is thus restricted to the expansion of F0 range and/or a move-up of the pitch register while the lexical tone contour remains intact (Lin, 2004b; Xu, 2017). In English, on the other hand, would assign a pitch target to make the focused domain salient (Pierrehumbert, 1980). Nevertheless the pitch range of the accentuation is larger than those without (Xu and Xu, 2005). Therefore it is safe to say that along the realisational dimension of LILt (Mennen, 2015), Chinese and English seem to bear a great number of similarities in terms of pitch accents, so it is not surprising to see such findings in which Chinese learners' production of accentuation exhibits similar properties to native English speakers', such as longer duration, louder intensity, and wider pitch span such as in Ji *et. al.* (2009). Chinese learners seem to be able to positively transfer these realisational strategies to the way they realise English.

With regards to the overuse of falling tones and underuse of rising tones on accented syllables, one plausible explanation is associated with the differences between Chinese and English on the systemic dimension of LILt (Mennen, 2015) in terms of pitch accent. Mandarin Chinese has lexically-based tones instead of the prosodically-based pitch accents of English. Thus it is not strictly accurate to compare Mandarin tone types to English pitch accent types, but in order to source the difficulties that Chinese learners were found to experience with particular English nuclear tones, it seems worth doing this comparison as in Liu and Chen (2016). By definition, the systemic dimension in LILt refers to the inventory of intonational categories (Mennen, 2015). As described in section 2.2.1, the four phonemic Mandarin Chinese tones are Tone 1—high level (HH), Tone 2—mid-rise (MH), Tone 3—low dip (LL), and Tone4—high fall (HL), so there seems no equivalent tone pattern for English low rising tone (L*H) as most English native speakers use in yes-no questions in Ji *et al.* (2009), but a similar tone pattern can be found as in Tone 3. According to the prediction of SLM (Flege, 1995) for L2 segmental phonology, if the L2 category is phonetically similar to a particular L1 category, the formation of a new category for the L2 will fail and learners will be faced with difficulty perceiving and as a result properly producing this L2 category. So if the prediction is expanded to L2 intonation, the English low rising tone, a typical nuclear tone is presumably difficult for MC learners (see Liu and Chen, 2016). However, this is not necessarily a legitimate answer to the failure of production of this type of pitch accent by MC learners in the studies abovementioned (Ji *et al.*, 2009; Shao *et al.*, 2011; Wang *et al.*, 2011a), because the method they used to elicit learners' production involved not just phonetic and phonological ability but also semantic and pragmatic knowledge of English intonation. It is by no means clear whether the participants were unable to utter low-rising tone or just unaware of using this type of pattern to signal questions and non-finality.

Therefore, studies which have attempted to examine the realisational and systemic dimension of intonation should carefully control for the meaning or contextual access for speakers (Gili Fivela, 2012). The following two studies have used rational methodological designs and

achieved rigorous results that MC learners of English do have difficulty in producing particular nuclear tones. Liu and Chen (2016) elicited twelve Mandarin speakers' productions of a pair of nonsense word DAga and daGA which merely differed in lexical stress (capitalised syllable). The two target words were carried by some simple declarative and interrogative sentences and placed at either middle or final position of these sentences. The qualitative and acoustic analyses of the production of both words in all conditions (sentence type × position) revealed that Chinese learners only projected L* pitch target on daGA when it was in sentence-final position of interrogatives. The failure of producing L* in any other positions (in DAga and the middle of sentences) led to a conclusion that Chinese learners seemed unable to systematically produce the low rising pitch accent of English due to the difference of systemic and realisational dimensions between Mandarin and English. The exhibition of L* on the final syllable of interrogatives was thought to influence the boundary H% tone of Mandarin Chinese for questions (Liu and Chen, 2016). The findings from the other study seem to more straightforwardly indicate the difficulty of producing English nuclear tone patterns. Zhou *et al.* (2012) directly tapped into the production of four English tones (high fall, high rise, fall-rise, and complex level) on the non-English word (*ma*), one/two-syllable English words and short sentences without referring to any contextual meanings, and found that the production of high fall was the closest to the native norm while the production of fall-rise was the worst. The phonetic context seems to play a role in the acquisition rate of different tone patterns too, as in sentence condition, the production of all four tones were the most deviant from the native norms, while in the non-word monosyllabic situation, the production of four tones was the best. This suggests that the Mandarin lexically-dominant tone system seems to exert an impact on learners' production of English tone patterns (Zhou *et al.*, 2012). If the English tones are applied to single syllables, it will be easier for them to be produced appropriately, but if the tones are hooked beyond the domain of the accented syllable (i.e. the tail in an intonation unit from the British approach), it will very likely become difficult to produce them.

When examining the differences of the phonetic realisations of de-accentuation, MC learners apparently assigned higher F0 and greater intensity, similar or longer vowel duration than native speakers did (Chen *et al.*, 2001; Chen, 2008a). The distinctions between stressed and unstressed syllables in terms of their duration was the least obvious from native speakers of Beijing Mandarin compared to Hong Kong Cantonese speakers (Chen, 2015b). Taking vowel quality into account, native speakers of Cantonese were found to be incapable of weakening syllables when speaking English, as they tended to use full vowels rather than schwa or syllabic consonants (Setter, 2006). This has been naturally attributed to the influence of L1 as both Cantonese and Mandarin are tone languages and prosody serves as an alternative to express semantic and pragmatic meanings only if syntactic cues are absent for in fact lexical tone-bearing syllables leave little room for phonetic flexibility (Chan and Li, 2000; Lin, 2007). Another explanation, provided by Deterding (2010), is far beyond L1 influence. He observed that MC learners tend to have a deeply-rooted belief that complete production of individual words is right and it is lazy to reduce vowels in speech; this is partially due to the extensive rote training of memorising vocabulary lists. Words with schwas in citation forms are not difficulty for them to produce, but centralised vowels in function words seem to be doomed in their production. Wang *et al.* (2010b) and Chen (2008a) have provided supportive evidence to Deterding (2010) as both have discovered that the stressed or accented syllables in Chinese production of English are almost equally distributed on content words and function words; improper placement of accentuation accounted for 48% in Chen's (2008a) study, particularly on pronouns which rarely attract stress unless it is in contrastive or informational focus in native speech. In addition, it was found that learners' production of English sentences seems to be in a multiple-accent fashion, by which the longer the sentences are, the more accents show up, and the number of accents within an intonation phrase turns out to be dramatically more than native production (Schack, 2000; Chen, 2008a; Wang *et al.*, 2010b). Another important aspect concerned with de-accentuation entails the PFC (post-focal compression). Both English (Xu and Xu, 2005) and Mandarin Chinese (Wang and Xu, 2011)

show evidence of a compressed and lowered pitch range in the post-focal position of the utterances. However interestingly, acoustic analyses have found that MC learners' production of English seems to be absent from PFC, suggesting that it might not be transferrable in this regard (Xu, 2012).

As for phrasing, the deviances have been found for boundary positioning and the phonetic and acoustic manipulation of boundary markers. Pause is a typical boundary marker in both Mandarin (Cao, 2012) and English (Cruttenden, 1997). Chinese learners tend to rely on pause to cut their utterances, and the duration of their boundary pauses seem significantly longer than those produced by native speakers (Chen, 2006; Li *et al.*, 2006), and the frequency of pausing by learners is greater than native speakers' (Anderson-Hsieh and Venkatagiri, 1994; Chen, 2008a). Since the sampled speech was all read speech in the studies reviewed so far, Timkova (2001) explains that for L2 learners, they tend to be oriented towards language itself (word identification or pronunciation) in the reading process during which they need more time to move from one word to another in processing, while their focus is hardly on intonation. But unfortunately, spontaneous speech in intonation research seems very rare to the author's knowledge due to methodological problems such as the complexity in transcribing the intonation and the arbitrariness of productions. Other prosodic markers such as anacrusis and pitch reset have been found rare in English by Chinese learners (Chen, 2006; Chen, 2008a), neither of which can find their counterparts in the realisation of prosodic boundaries in Mandarin Chinese (see section 2.2.2). These difficulties can be reasonably classified into the realisational dimension of LILt (Mennen, 2015). Although Mandarin Chinese has both L% and H% as the boundary tone (Lin, 2004a), systemically similar to English boundary tones (Pierrehumbert, 1980; Ladd, 2008), they are principally used for distinguishing questions from statements which is semantically constrained compared to their usage in English. So it is anticipated that for English H%, which is also used for the situation of non-finality and continuation, will be difficult for MC learners to acquire. While difficulties that have been found in the positioning of boundaries, learners' performance

seems to be more chaotic and random as shown in Chen (2008). For instance, some learners paused in between verb phrases, compound nouns, and noun phrases, and made final reporting phrases as an independent intonational phrase from its main clause. Some even isolated the relative pronouns (e.g. *that*, *which*, etc.) from the clauses, all of which made their speech less coherent and intelligible. These difficulties can be considered in the semantic dimension of LILt (Mennen, 2015), but one should not jump to the conclusion that learners cannot produce proper phrasing because their L1 lacks equivalent semantic/pragmatic use of intonation phrase, and in fact, Chinese barely treats syllables within prosodic words (compounds, verb phrases, etc.) separately with noticeable boundaries (see Cao, 2012). What makes Chinese learners do all these has yet to be made clear and needs to be examined by means of stratifying different levels of ability as specified in the LILt.

2.3.3 L2 learners' perception of intonation

The LILt (Mennen, 2015) posits that L2 learners' difficulties in the production of intonation seems to be perception-oriented, as predicted for the L2 phoneme acquisition by theories such as SLM and PAM-L2 both of which are bound to the learners' quality of perceptual discrimination of L2 segmental categories (Gili Fivela, 2012). However, categorical perception of L2 intonation is under-examined. One of the major reasons is that, unlike segmental categories, intonation categories arguably "have very general meanings and play specific functions at the sentence level" (ibid: 20). So from a technical point of view, it requires an additional consideration of contextual meanings corresponding to the intonation categories that are about to be investigated. Such extra effort has made those traditional discrimination methods that are applied to segmental levels less suitable for intonation studies (Massaro, 1998). For example, in studies such as Liu and Rodriguez (2012), sentence-final syllabic F0 offset was synthesized to the values along a continuum between 200Hz to 300Hz with the F0 onset fixed at 250Hz. L2 learners were asked to identify whether the contour they heard was a question or statement and to discriminate if the two stimuli were the same or

different. The anomalous results for the discrimination function (i.e. modest peaks) to some extent confirmed of Gili Fivela's (2012) argument that categorical perception of intonation by such a method lacks precise definition of intonation functions and eventually causes ambiguity in qualifying assimilation types. Another study (Sánchez-Alvarado and Armstrong, 2016) which manipulated the pitch scales of the L+H* pitch accent in Spanish tested the hypothesis that native Spanish listeners would perceive contrastive focus within a more compressed pitch range than American learners of Spanish, because both languages have this type of pitch accent but the phonetic realisation of Spanish has lesser pitch scales than its English counterpart. But the results did not support their hypothesis as both groups performed better at the higher end of the scale, leading to a suspicion that non-focal phonetic realisation might also play a role in the perception of contrastive focus (Sánchez-Alvarado and Armstrong, 2016).

Therefore, instead of tapping into the categorical perception of intonation along a particular phonetic dimension that is irrelevant to or loosely connected to meanings, studies in this area have turned their attention to those naturally produced stimuli that are specified for phonological and pragmatic functions of intonation. The findings from such studies contribute not only to complementing the theories of L2 intonation acquisition, but also enlightening the teaching of intonation. The following studies seem to be in this vein. He and his colleagues investigated the extent to which MC learners of Dutch can make correct choices of pitch contours to contextualised sentences by comparing them to native performance (He *et al.*, 2012b). They selected 26 sentences from a story, presented to the subjects in a chronological order. Each sentence was provided with four intonation contours auditorily and visually, one of which was thought to be the optimal choice while the other three were distractors. They argued that if Chinese learners cannot match the optimal pitch contour to the highly contextualised sentence, their inability of producing correct prosody might be due to the lack of relevant semantic and pragmatic knowledge of intonation rather than their inability of perceiving acoustic cues. The results showed that Chinese learners did significantly worse

than native participants (53.3% vs. 83.3% correct rate), though more-advanced students did significantly better than less advanced students (61% vs. 52% correct rate). The Entropy measures revealed that Chinese listeners had significantly less agreement over what was the best and the worst choice compared to native listeners, and the lower proficient learners had a lower degree of agreement than more advanced learners. This suggests that L2 learners did experience difficulty in the perception of intonational meanings. Following He *et al.* (2012b), Mok *et al.* (2016) conducted a similar study looking into L1-Cantonese L2-English Hong Kong listeners' selections of the most appropriate intonation contour for sentences extracted from a story. The target sentences covered a wide variety of sentence types and each sentence was generated with five English nuclear tones (fall, rise, fall-rise, rise-fall, and level) on the same nuclear word. They found that Cantonese listeners did worse than the natives in general, and the L1 influence was discovered for certain sentence types that lack Cantonese equivalents such as tag questions. Learners had difficulty choosing the optimal contour, while those with Cantonese equivalents gained higher accuracy rates. Puga *et al.* (2017) then replicated Mok *et al.*'s (2016) study but targeted at German learners of English, hypothesizing that German learners would do better in this type of tasks for German is typologically closer to English as both are intonation languages. The results supported the L1 influence that had been found in Mok *et al.* (2016); German learners did similarly well as native English listeners did in sentence types that both German and English have, such as statement questions, echo questions, continuation and yes/no questions. Grammatical structures like open/closed tags and checking tags absent in German turned out to be challenging. Both studies, however, agreed that L1 influence alone cannot fully explain learners' performance because: 1. syntactic structures existing in L1 and L2 imposed challenge, e.g. wh-question for Cantonese learners, sarcasm for German learners; 2. German and Cantonese learners exhibited similar difficulties in certain structures regardless of their first language such as tag questions. This has led to a conclusion by Puga *et al.* (2017) that there might be some universal perception errors across L2 learners. It should be noted that in both studies, native

listeners did not perform at the ceiling level for particular items such as echo questions, checking, yes/no questions, indicating that intonation contours associated to specific context/meanings are in fact far from consensus among native speakers. Thus the teaching and learning of these form-function relationships needs to be cautious, while structures like wh-questions and tag questions which attract unanimous choices over the ideal contours among the native speakers might deserve emphasis in teaching.

As for the perception of prominence, the phonetic cues independent of contextual meanings have been sought by Rosenberg *et al.* (2010) in which sentences composed of four monosyllabic or disyllabic words with different parts of speech (determiner, noun, verb, adverbs) were produced with three types of nuclear accents (H*L-L%, H*H-H%, L*H-H%) on each of the words. The word order was randomised as well except for the determiner which was always positioned before the noun. Mandarin EFL learners were asked to note down the prominent words while listening to the sentences. They found that the word length, word position and contour type had significant effect on learners' judgement of the position of prominence. Learners were better at identifying prominence when the pitch change happened on disyllabic adverbs and determiners which were placed at sentence-final position. The authors also measured and compared the acoustic qualities of words identified as prominent to those not and found that MC listeners tended to adopt similar cues as native listeners did to identify prominence, which seems to be in accordance with their production performance on prominence as reviewed before (e.g. Shao *et al.*, 2011). However, nuclear contours with H* target were not always easy for MC learners to choose the "correct" position of prominence, because H*H-H% (high-rising tone) was the least identified prominence, although H*L-L% (falling tone) was the best and better than L*H-H% (low-rising tone). This does not apparently match the findings for the production of H* and L* pitch accents in which H* was always the phonetic strategy for signalling prominence while L* was very rare because of L1 transfer (e.g. Ji *et al.*, 2009). So the hypothesis that the deviances in production is likely to be perception-based seems to be rejected in terms of pitch accents' realisational dimension of

LILt. The mismatch in the perception-production relationship has also been found in phonological and semantic dimension, as in the same study (Rosenberg *et al.*, 2010), Chinese learners were better at the identification of accented adverbs and determiners than accented nouns and verbs. It was explained that accentuation on some parts of speech like adverbs and determiners are less expected than accentuation on nouns and verbs, and this resulted in immediately noticing of those parts of speech when they were protruded. But studies on learners' production (e.g. Chen, 2008a) have found that Chinese have no preference for particular parts of speech as they tend to emphasize nearly all words, which means that if learners' production is indeed perception-oriented, there should not be any significantly favoured parts of speech in their perception of prominence.

Mixed results concerned with the perception of prominence have been found when different levels of processing are involved. The study by Wang *et al.* (2010a) only involved phonetic and phonological level of processing in which Chinese learners exhibited much more random and scattered perception of the position of English prominence than native listeners. The way they collected their data was listeners noting down the perceived prominent words in the sentences they heard which were produced by native speakers in the form of narrow vs. broad focus and statement vs. question. However, in another study (Gananathan *et al.*, 2015), Cantonese listeners were able to identify correct prominence position in a focus-inducing question. The latter study seems to be involved with semantic and pragmatic levels of processing by which a richer context is associated with prominence perception. Ortega-Llebaria and Colantoni (2014) directly compared L2 learners' performance on prominence for different levels of contextual availability, and they found that learners did perform differently on the meaning-free tasks from the meaning-rich tasks. Nevertheless the meaning-rich tasks seem to impose much more difficulty than the meaning-free tasks, which is in the opposite of the results from the two studies reviewed above.

The relationship between the perception and production of prominence has been in fact directly examined in these three studies. In Ortega-Llebaria and Colantoni (2014), both

perception and production tasks were divided into meaning-free and meaning-rich context. In the meaning-free context, learners just needed to listen to some unrelated sentences with different locations of prominence and then read out those sentences by mimicking the intonation patterns. For the meaning-free perception task, they again listened to the target sentence, and then listened to three low-pass filtered intonation (without identifiable segments) differing in the placement of prominence and picked out the one they just heard for the target sentence. For the meaning-rich production task, learners were told a story and asked 15 questions deduced from the story, they needed to produce an answer to each question by placing focus on the correct position. For example, the question would be like: Who is the dog? Is ***BOBBY*** the dog? The expected answer was: No, ***TOBY*** is the dog. The perception task for the meaning-rich context was the selection of the most appropriate patterns for the focus-eliciting questions, e.g. Question: Did Bobby fall out of the tree? Answers: (a) ***TOBY*** fell out of the tree; (b) Toby ***FELL OUT*** of the tree; (c) Toby fell out of the tree (Extracted from Ortega-Llebaria and Colantoni, 2014: 340). Participants in this study included L1-Mandarin and L1-Spanish learners of English and it was hypothesized that the Mandarin learners would perform better than the Spanish learners in the perception and production of English prominence because Mandarin uses similar prosodic manipulation for the focused words while Spanish normally adopts syntactic rather than prosodic means to the focusing function. The results have shown that L1 transfer was more evident in production than in perception, and L1 transfer increased with access to meaning, especially in production. In Gananathan *et al.* (2015), the perception outperformed the production, as Cantonese learners experienced little difficulty in the selection of the best prominence location but when they were required to answer the questions with proper prominence, the phonetic realisations deviated from the native norms. However in Wang *et al.*'s (2010) study, which was the least context-based, the perception and production of prominence were equally poor compared to native performance, indicating a closer relationship between perception and production.

Less is known about the perception of L2 intonation phrasing. Phrasing, in fact, is inherently bound with accentuation and possibly tone as well, meaning that the change of phrasing is often accompanied by an addition/deletion or shifting of accentuation. For instance, if //*The flags are red, white and blue.*// is treated as only one intonation phrase, the nucleus is on *blue* and it means the flags are all three-coloured. If it is uttered with three intonation phrases, as in //*The flags are red,// white// and blue.*//, it means some flags are red, some are white and the rest are blue. In each phrase, the colour is accented. This is possibly why phrasing has been rarely researched on its own from a methodological perspective, and it can hardly be disentangled from other elements of intonation. A handful of existing studies which integrated phrasing into investigation pertain to a higher level of processing—interpretation of intonational meanings of contrastive minimal pairs. For example, Cruz-Ferreira (1987) recruited 30 L1-English L2-Portuguese and 30 L1-Portuguese L2-English participants to explore the cross-linguistic discrimination and interpretation of non-native intonation patterns based on the British nuclear approach. Intonation minimal pairs were those with contrastive semantic or pragmatic meanings using by tonicity, tonality or tone. Participants were asked to discriminate if the two intonation patterns of the same sentence were phonetically the same or different and to select the correct meaning from two meaning glosses to the particular intonation pattern they heard. Three interpretative strategies for non-native intonation were proposed to account for the correctness and errors made by the two groups of L2 learners. The correct interpretation of intonational meanings might occur by means of the positive transfer strategy when the L1 and L2 use similar intonation forms to distinguish similar meanings, or by means of the pitch height strategy when the L1 and L2 use different intonation contrasts to deliver some universal meaning contrasts, e.g. higher pitch associated with openness and lower pitch with finality (Cruz-Ferreira, 1987: 115). The wrong or random interpretation of intonational meanings was found when the negative transfer or a lexico-syntactic strategy was in play. If the intonation patterns exist and contrast in both languages but are used to distinguish different meanings, learners are unlikely to map the meaning to the form in L2

correctly. This is the case when the L1 negatively interferes with the L2. If the use of intonation is idiosyncratic in the L2, the rate of correct discrimination and interpretation will be low and learners will tend to ignore the intonation and interpret the meaning literally (lexico-syntactic strategy) (Cruz-Ferreira, 1987). The proposed three interpretive strategies imply an intricate relation of intonation patterns and lexico-syntactic structures in the construction of intonation meanings and provide an insight into the prediction of L2 learners' behaviour in the comprehension of intonation. Atoye (2005) conducted a study with a similar methodological design to Cruz-Ferreira (1987), examining into 120 Nigerian learners' perception and interpretation of English intonation meaning contrasts by 1. listening to paired intonation patterns and deciding if they were the same or different; 2. interpreting the meaning of the pattern they heard for the target sentence. The results show that the subjects were able to discriminate the phonetic differences between intonation contrasts (with 85.7% correct rate), but their understanding and interpretation of intonation meanings were far from the "correct" one from the textbook (with 25.7% correct rate), albeit there was an agreement of interpretation among the subjects due to L1 transfer or influence. Both studies suggest that L2 learners are confronted with difficulty understanding the intonation meanings. The research into this aspect can be supplementary to the understanding of L2 acquisition of intonation, but the documented literature so far seems to be too limited to make itself heard.

2.3.4 Other issues related to L2 acquisition of intonation

The most popular and widely researched topic is the correlation of L2 experience and intonation competence. A rich body of research which focuses on the comparison of learners' production and perception of intonational features to the native norms has consistently found that students with more advanced L2 proficiency tend to exhibit more native-like performance and those with lower levels of proficiency seem to be more vulnerable to the L1 influence (e.g. Anderson-Hsieh and Venkatagiri, 1994; Li *et al.*, 2006; He *et al.*, 2010; Wang *et al.*, 2010a; Wang *et al.*, 2011a; He *et al.*, 2012a). However, when learners' production was

listened to and evaluated by native listeners on the degree of foreign accentedness, some intonational features such as pause duration and frequency were not correlated with L2 experience. Learners with 10 years of learning experience did not show any significant advantages in these aspects than those with only 3 months or 3 years of experience (Trofimovich and Baker, 2006). On the other hand, empirical research on higher level of intonation processing, e.g. interpretation of intonation meanings, to the researcher's knowledge, has yet compared different levels of L2 proficiency. So it is still inconclusive as to whether the increase of L2 experience leads to proportional improvement of intonation competence at all levels or dimensions of L2 intonation.

Nonetheless, the evidence that has been reported for the positive correlation between intonation competence and L2 experience has lent support to the argument that intonation is learnable over time either with or without intervention. One direct reference arises from a longitudinal study which explores the natural development of L2 learners' perception of five pronunciation aspects: word stress, sentence stress, intonation, can/can't distinction, and numbers (13 vs. 30) (Derwing *et al.*, 2012). Over a 10-month period of time, Mandarin and Slavic learners of English showed improvement on the perception of sentence stress and intonation without explicit instruction, although intonation contours used here were restricted to rising tone and monotone. In Zhou *et al.*'s (2012) study, Chinese learners improved the rate of correctly producing four English tones including the falling-rising tone that was thought to be the hardest for this group of learners under 18 hours of instruction. Both suggested that intonation deserves to be taught for the purpose of communication. Even if some intonational features or functions can be perceived naturally without intervention as reported in Derwing *et al.* (2012), it does not necessarily mean that production will develop in parallel with perception and needs no instruction. And in fact, many researchers have acknowledged that native-like production of intonation contours may just be a reflection of imitation skill (Cruz-Ferreira, 1987). Instruction should therefore be based on the semantic and pragmatic knowledge of intonation, as the failure of mastering such knowledge impedes learners'

understanding of intonation meaning and causes the erroneous uses of intonation forms (Toivanen, 2003; Atoye, 2005; He *et al.*, 2012b).

Universality has often been brought up as an alternative explanation of L2 learners' common or similar production on particular intonation features across diverse L1 backgrounds. For example, L1-German Austrian learners of English acquired simple nuclear tones easier than complex tones and the falling tone seemed to be always their preferred tone assignment when producing English, as even with one year of instruction their adoption of rising tone accounted for only 18% of the tone choices (Grosser, 1997). L1-Mandarin Chinese learners have shown a similar acquisition order by which falling precedes rising and falling-rising tone (Zhou *et al.*, 2012). In the latter case, the failure of grasping the rising and falling-rising tone was ascribed to the markedness of intonational features, as the falling tone is less marked than the rising and falling-rising tone (Zhou *et al.*, 2012). The Austrian learners in the former case used rising tone for 46% of the time in their native language of German, contradicting their rare usage of rising tone in English (Grosser, 1997). This offers a rethinking of the underlying causes of the difficulties for a particular group of L2 learners; L1 transfer plays a crucial role but universal markedness is surely playing a role as well. Eckman's (1977) *Markedness Differential Hypothesis* argues that some language structures which are difficult per se are marked cross-linguistically. Rising tone, as a marked pitch pattern, poses challenges to foreign language learners, even though some languages use rising tone very often, as in Grosser (1997), the learners failed to transfer the L1 (German) use to the L2 (English). A broader picture is provided by cross-linguistic investigation into L2 learners' similarities and dissimilarities of their intonation competence. Mennen *et al.* (2010) has characterised the internal structure of intonation along systemic, realisational, semantic and frequency dimensions by L1 Punjabi and Italian learners of English who had just arrived at the English-speaking country. Although Punjabi and Italian are typologically distant in terms of their prosodic systems, learners from the two countries were found sharing the same pitch accent and boundary tone inventories, employing similar acoustic manipulations of accentuations,

grouping similar number of accented words into their intonation units. Hewings (1995) also found that learners with L1-Korean, L1-Greek, and L1-Indonesian background were all incapable of signalling discoursal meanings with rising boundary tones such as in agreement withholding, reservation, and contradiction. The evidence presented in this section has clearly implied that transfer seems not the only source of L2 prosodic difficulty. The assertions of L1 influence need to be cautiously made as many other factors apparently play a role in L2 acquisition of intonation.

2.3.5 Summary

This section has reviewed the existing literature on L2 learners' acquisition of intonation, particularly targeting Chinese learners of English for whom are witnessed various but repeatedly reported difficulties in both perceiving and producing English intonation on multiple dimensions. It has also given a detailed review of the recently proposed L2 Intonation Learning theory (LILt) (Mennen, 2015) under which learners' production difficulties can be modelled into four intonational dimensions that allow systematic comparisons across languages. Moreover, it has identified issues that have not been satisfactorily addressed and left open to further empirically-driven research. Firstly, the underlying causes of learners' poor performance seems far from clear. As recognised by Mennen (2015), for instance, the deviance in the realisational dimension of intonation might stem from the difficulty with the systemic dimension, and vice versa—the absence of particular phonological categories does not necessarily mean that learners do not have this type of category in their interlanguage inventory. It might be a result of insufficient knowledge of intonational semantics that correlates to the usage of certain patterns, or due to the distorted phonetic realisations of a particular type of category.

Secondly, the perception of intonational meanings by L2 learners is under-researched, resulting in unwarranted pedagogical implications that teaching intonational form and structures is enough for polished production of native-like pitch patterns. Nigerian speakers of

English turned out to be unable to interpret the correct meanings of English intonation contours even though they were trained for intonation with structural approach (Atoye, 2005). Thus it is anticipated that learners who cannot understand intonational meanings will encounter difficulty in real-life conversational situations. It is therefore reasonable to explore first if L2 learners can or cannot understand intonation functions and meanings carried by intonation forms that are encrypted with multiple phonetic dimensions. If they cannot, we can source their difficulties at the phonetic and phonological levels to the lack of knowledge of the semantics and pragmatics of intonation and the relation of intonational form and functions.

Last but not least, there is little research on L2 learners' learning path of intonational elements, i.e. tonicity (accentuation), tonality (phrasing), and tone. Production studies have found an interrelated effect of learners' difficulties in different compositions of intonation—pausing errors and accentuation errors are significantly correlated (Rasier and Hiligsmann, 2007). Chen (2013) found that L1-Mandarin Chinese learners' difficulty in rising tone is associated with improper placement of nuclear stress, failure to control the nuclear contours, substitution with falling tone, and epenthesis of schwa following the phrase-final consonants. And he further points out that the acquisition of rising tone pertains not only to tone itself, but also to the acquisition of sentence stress. These interesting findings seem to reveal that L2 learners' acquisition of intonation elements influence each other and some elements might be more difficult to acquire than others. Research on L1 child prosodic development might shed light on this issue, as between the age of one and four, children have started to use contrastive tones to distinguish sentence types and functions. Accentuation has also started to appear more or less simultaneously with tonal contrasts. Chunking comes in later only when children begin to produce longer sentences (Crystal, 1986). While children's comprehension of intonation meanings delivered by these elements seems to follow this route, the comprehension of tonal contrasts and tonicity placement develops earlier than the comprehension of tonality (Cruttenden, 1985). It is worthwhile to explore if L2 learners'

production and perception of intonation elements exhibit a similar fashion in which phrasing is more difficult than accentuation and tone.

2.4 Computer-assisted pronunciation teaching for intonation

As early as the 1990s, when pronunciation was a resurgent interest to teachers and researchers, computer-assisted pronunciation teaching (CAPT) started to be highly valued for multiple reasons. First, it offers individualised instruction and feedback on learner performance. Second, there are no limits accessing L2 speech input either inside or outside the classroom. Third, it is full of a great deal of exercise for L2 speech perception and production. All of these reasons are presumably legitimate for a successful mastery of L2 pronunciation (cf. Levis, 2007; Felps *et al.*, 2009). It is especially so for Chinese learners' situations where most of their class time is occupied with linguistic skills other than pronunciation, and whenever pronunciation is instructed, teachers are always criticised for ineligibility in L2 phonetics and phonology. Finally individual learner differences are rarely considered due to the large size of class (Lu, 2010a; Shi, 2010).

The core endorsement for CAPT is its integration of automatic speech recognition (ASR) and electronic visual feedback technologies (Anderson-Hsieh, 1994; Chun, 1998; Levis, 2007), whereby learners can get instant real-time feedback on the degree of nativeness of their speech or comparable visualisation of speech signals such as waveforms, spectrograms, and pitch curves between native samples and thier own production (Hincks, 2003; Hardison, 2004; Ding *et al.*, 2012a; Nicolaidis *et al.*, 2015). The former type of feedback (ASR) is usually programmed into commercial CAPT tools such as Tell Me More (by Auralog). Learners' production of L2 speech is recognised and examined by a probability-based model, i.e. Hidden Markov Model (HMM), through which the level of pronunciation accuracy is identified by statistical comparisons to the large database of native samples stored in the system (Godwin-Jones, 2009). So inevitably one of the biggest challenges for ASR-based

feedback seems to be the inaccuracy in capturing non-native errors as the ASR is trained by native speech and it is sensitive to and ideal for the recognition of native voice. As revealed by Ehsani and Knodt (1998) the accuracy rate of some ASR-based tools for native users reaches up to more than 95%, while being used by proficient but accented non-native users, it decreases drastically to 70% (Coniam, 1999). And it is particularly difficult for targeting prosodic errors, "because all languages make use of the same acoustic categories of pitch, duration, and intensity for differently organised prosodic systems" (Levis, 2007: 193). Such recognition errors make the feedback unreliable and might cause detrimental effects during L2 pronunciation learning (Neri *et al.*, 2002). Another problem for ASR-based feedback seems to be the failure of demonstrating the remediation of learner errors. Students can hardly get explicit explanation of how to amend the errors and improve their pronunciation from such feedback (Engwall and Bälter, 2007).

The other type of feedback is associated with the visualisation of individual speech sounds or prosodic aspects of longer strings. Applied linguists seem to adore CAPT tools embedded with visualisation technologies vs. those with ASR as they proclaim that visualised cues "provide a secondary modality to facilitate noticing" when auditory perception is biased by learners' L1 (Olson, 2014: 185). Moreover, visualised acoustic entities of native speech have set up a standard for L2 learners' productions. By comparing their own production to the model, learners can identify and approach to their errors relatively easier than ASR-based feedback. The major issue with visualised feedback concerns the transparency of the interpretation of the visual displays; spectrograms and waveforms are mostly challenging (Levis, 2007), as the understanding of what these acoustic cues tell about articulatory controls needs specific instruction and it normally adds extra burdens on L2 learners (Setter and Jenkins, 2005). Pitch tracings, on the other hand, seem to be more interpretable—rising or falling pitch curves intuitively correspond to raising or lowering a speaker's tone of voice (Levis, 2007: 191). The recent psycholinguistic studies have evidently corroborated that "the mental representation of pitch is audiovisual in nature" (Yuan *et al.*, 2017: 3), and the

cognitive load of comparing the difference between pitch contours is significantly heavier with auditory cues than with visual cues (Hermes, 1998), leading to a reaffirmation that visualisation of pitch could aid the acquisition of intonation.

A great deal of empirical research on the effectiveness of such CAPT techniques has consistently reported the positive effects for L2 learners' acquisition of intonation. For instance, in Ramírez Verdugo (2006), twenty Spanish learners of English were randomly assigned to a treatment and a control group. The treatment group received instruction on intonational form and functions and trained with *Speech Analyzer* which visualised native samples as well as Spanish learners', whereas the control group had neither the instruction nor the audiovisual training during the period of the experiment. Pre- and post-test design was observed with a global improvement in the treatment group's intonational structure, e.g. tonic prominence was more clearly marked in intonation units in the read speech than that of learners from the control group. Hardison (2004) conducted instructor-monitored individual training on 16 American learners' production of French intonation by a costly commercial visualisation program called *Real-Time Pitch* from the Computerized Speech Lab of Kay Elemetrics. The training procedure lasted for three weeks and covered 13 sessions each of which was 40 minutes long. Participants were required to recite the training sentence and record it. Audiovisual feedback from the native production of that sentence was displayed on half of the computer screen, together with the visualisation of their own production on the other half of the screen. Then the native pitch curve was overlaid on the learner's in a different colour. When another rendition was being recorded, the previous production was erased. The improvement of overall nativeness was evaluated on a scale of 1-7 by native French speakers by listening to filtered and unfiltered learners' productions of French sentences. It was found that audio-visual F0 feedback not only led to an increase of nativeness of learners' intonation but also improved their production of segments, and this effect spilled over to the production of novel sentences. Similar positive effects on learners' production of

overall prosody were also found in Anderson-Hsieh (1992), Hardison (2005), Le and Brook (2011), and Gorjian *et al.* (2013), albeit with different pitch visualisation tools in use.

Unfortunately, none of the above-mentioned studies have included a control group of participants trained with auditory-only feedback, which seems to undermine the robustness of the argument that visualised pitch promotes perception and production of intonation. Visualised feedback referred in these studies are in fact audio-visual per se; learners are guided through the visual cues simultaneously with auditory recordings. The recent work on auditory training for L2 intonation has revealed that auditory feedback can facilitate the acquisition of intonation once the basic knowledge of intonation is explicitly imparted prior to the training. Tanner and Landon (2009) recruited 75 English learners with a wide range of L1 backgrounds and appointed them to the experiment and the control group. Both group received 65-minute intonation instruction specified for pausing, word/sentence stress, and final intonation pattern (rise vs. fall). Only the experiment group were required for additional 10-minute self-directed CAPT ear-training for homework on a daily basis over the following 11 weeks. The CAPT tool was called cued pronunciation recording (CPR) by which participants would be working on a passage written on the computer screen. While listening to the native speaker's reading out of this passage, participants marked the pauses, word and sentence stresses and sentence-final intonations based on the prior instruction. This process could be done multiple times before they were happy about their markers and recorded their own production of the passage according to the markers. Then they were able to access their own production and auditorily compared to that of native recordings and re-recorded themselves. By comparing participants' performance on a battery of perceptual and productive tests before and after the CPR training, it was found that learners in the experiment group exhibited obvious improvement in the perception of pause, word stress and sentence-final tone and the production of stress, though the production of pause and tone were not significantly improved. Notwithstanding the limited tangible evidence on improved production with auditory training, it suggests that the intonational feedback via auditory

modality alone can foster L2 learners' perceptual ability. If learners' production of L2 intonation is perception-oriented as predicted by the LILt (Mennen, 2015), the improvement of production with audio-visual training reported above might be partly due to the improved perception as a result of training outcome from the auditory feedback.

One existing empirical study, though published almost 35 years ago, attempted to address this issue by investigating the overall nativeness of Dutch learners' production of English intonation from three groups of treatments: instruction, instruction with auditory training, and instruction with audio-visual training compared to the control group who merely took the pre- and post-test. de Bot (1983) analysed the overall scores rated by native judges on a 5-point scale for all participants with pre- and post-test as within-subject variable, and treatments and practice time (45 vs. 90 minutes applied to the auditory and audio-visual group) as between-subject variables. The results were interesting as with an instruction of only 15 minutes long, learners' imitation of 15 test sentences was significantly better than they did in the pre-test. Probably due to the ceiling effect, the following practice for the other two treatment groups did not yield satisfactory effect, although the audio-visual group showed a slightly larger improvement than the instruction-only group. Training time did not play a crucial role in improving learners' imitation skill. Learning behaviour was also recorded in this study and provided some valuable insights for intonation training. de Bot (1983) found that the audio-visual group tended to repeat and practice more often than the auditory group who spent most of their time on listening to the models without oral practice. Thus the author drew the conclusion that "the most likely explanation of the positive effect of audio-visual FB (feedback) as found in this experiment is that subject are provided with additional (visual) information regarding specific errors in the subjects' imitations of the target sentences. In addition, the use of this kind of equipment tends to increase the subjects' motivation to try harder to correct an error" (ibid: 348). With very limited existence of empirical research such as de Bot's (1983), it is by no means convincing that audio-visual feedback is more efficient than auditory feedback in L2 intonation training. It should also be noted that learners'

intonational competence tested and trained throughout the experiment was focused on imitation skill which is outdated from a pedagogical point of view. This actually mirrors the primary drawback of most CAPT applications that have been heavily criticised for—the obsession with technological novelty over pedagogy (Levy, 1997; Neri *et al.*, 2002; Martin, 2004; Clifford and Granoien, 2007). Levis (2007: 185) metaphorically referred it to "traditional, drill-oriented pedagogy in new clothing".

Trouvain *et al.* (2016) evaluated the performance of ten CAPT programs on the market within a framework which is pedagogically grounded and list their collective shortcomings into six categories: 1. Learners' L1 and L2 proficiency levels are rarely taken into consideration. 2. Instructions suffer from lack of clarification. 3. The coverage of pronunciation aspects is not wide enough as prosody is frequently excluded. 4. Metalinguistic feedback on errors is seldom provided. 5. The learning progress is often not controlled. 6. The learning goals are rarely reflected in exercises. Apart from these pedagogical drawbacks, a more serious issue particularly associated with prosodic features seems to be insufficient theoretical descriptions of intonation functions. CAPT applications, such as *Streaming Speech* (Cauldwell, 2002) that is strictly based on the discoursal functions of intonation, are extremely rare (Levis, 2007).[8] In addition, most of the CAPT programs are commercially pricy. If the school declines the sponsorship of such programs, teachers will have no choice but to offer little feedback to individual students in large pronunciation classes by themselves (Luo, 2016).

All in all, although the existing CAPT applications have been undergoing criticisms from both technological and pedagogical sides, the hopes for using them by researchers and language teachers seem to remain high as they have potential benefits that traditional pronunciation classes cannot compare to (see details from the beginning of this section). More importantly, students are prone to be attracted to and motivated by a pronunciation course facilitated by computer (Lu, 2010a), and tend to behave more lively and show more

[8] *Streaming Speech* is a commercial CD-ROM application for intonation learning. It is theoretically based on Brazil's (1980, 1994) Discourse Intonation.

engagement in learning (Gorjian *et al.*, 2013). Individualised interaction with the computer can also minimize anxiety and embarrassment that students might experience in a traditional class (Murray, 1999). Practical use of these applications, however, is often held back because of the gap between the contents in the CAPT tools and the pedagogical requirements. Prosodic or intonational elements have been called for priority in teaching and learning, but applications specialised in these areas are unfortunately far less satisfactorily developed. Investigations of the effectiveness of the CAPT tools for L2 prosody are usually conducted with the least involvement of instruction (e.g. de Bot, 1983; Hardison, 2005; Wilson, 2008; Hincks and Edlund, 2009; Le and Brook, 2011) as if classroom teachers can be substituted, the idea that has been unreservedly objected to (Setter and Jenkins, 2005; Trouvain *et al.*, 2016). Saito (2012) reviews 15 previously published studies on the effect of pronunciation instruction (on segments and suprasegmentals) and finds that positive impact on subjects' production was consistently reported in 13 papers, while the remaining two in which no such effect was found was either due to the ceiling effect or to the short length of instruction (less than 30 minutes). So it seems safe to draw a conclusion that the teaching and learning of L2 intonation could be open to CAPT tools, but it requires the teacher to have a wise judgement of the most suitable tools for a reflection of his/her pedagogical needs. Speech visualisation tools, might be a good choice based on the literature review.

Chapter III: Methodology

This chapter is organised into four sections. The first section puts forward the research questions derived from the literature review in Chapter II. The second section describes the demographic information of Chinese participants and presents a robust analysis of their English proficiency, followed by a detailed depiction of the tailor-made instruction materials and practice materials for the training phase in this experiment in section three. The fourth section describes how the intonation training was delivered for the instruction and the practice.

3.1 Research questions

Enlightened by the review in Chapter II, the present study chooses EFL Chinese learners in the UK in English-related majors as the experimental targets because some of them will become English teachers back in China after achieving their degrees. I argue that training these potential teachers can be served as the first promising step towards the fulfilment of the *Guidelines on College English Teaching* (GCET, 2016). Only with enough qualified pronunciation teachers can the goals of cultivating competitive students become approachable.

Therefore, the primary aim of this study is to raise Chinese future English teachers' awareness of the importance of intonation by showing them how the changes of intonation patterns can lead to a significant change of the meanings, and by demonstrating the advantages and disadvantages of learning English intonation drawing on the findings from the relevant empirical research. The secondary aim is to help improve their intonation performance with the aid of computer-assisted practice supported by theoretical and pedagogical evidence. *Praat* and *Audacity* are two appropriate computerised tools to aid practicing of English intonation, a learning process that the participants can control for by themselves. *Praat* is a

popular computer application for speech analyses. There are three-fold reasons for choosing *Praat* (Boersma and Weenink, 2009*)*. First, it is free and downloadable for Windows and Macintosh. Second, it records speech and visualises pitch of pre-prepared sound files and the recordings. Last and most crucially, it leaves teachers with great flexibility in designing practice materials that can be arranged and displayed according to pedagogical considerations. *Audacity* (Version 2.0.0) is a speech editing application; it cannot show pitch tracings but captures recordings and saves them for further listening. Students using *Audacity* as a learning tool will be "blind" to intonation. All they can access will be the audio recordings of pre-prepared native talks and their own production. The effect of audio-visual feedback on intonation acquisition has not yet been uncontroversially substantiated due to the failure of an inclusion of a controlled auditory group for comparison. So, the last aim of this study is to compare the effectiveness of audio-visual and auditory feedback on Chinese EFL learners' intonation acquisition. A control group who is not involved in any training process will also be included in order to rule out the possible natural acquisition of intonation that contributes to the improvement that might be found for the training groups.

Participants' intonation performance was assessed by means of a comprehension task as it is the least researched intonation ability of L2 learners as reviewed in section 2.3.3 and 2.3.5. It is argued that without the knowledge of learners' perception of intonation meanings, there is no way to source the underlying causes of their difficulties that have been reported at the phonetic and phonological dimensions (e.g. Chen, 2008; Ji *et al*., 2009; Wang *et al*., 2011a). Little research has tapped into L2 learners' perception of intonation meaning (see section 2.3.3 where Cruz-Ferreira, 1987 and Atoye, 2005 are reviewed). Thus, the first research question that this thesis attempts to address is can Chinese EFL learners distinguish the meanings conveyed by contrastive intonation patterns in terms of tonality, tonicity and tone, and is there any difference between their identification of tonality, tonicity and tone?

Previous research has discovered the effectiveness of intonation training on L1-Mandarin Chinese learners' production of British nuclear tones (Zhou *et al*., 2012) and the overall pitch

contours associated with pragmatic meaning (Gao, 2015), but it is not clear if intonation training is effective for improving learners' perception of intonation meaning. So the second research question is, with explicit instruction on English intonation and learner-paced additional practice, will their identification of intonation meaning be improved? If so, is there any lasting effect of this kind of training with a pre-, post-, and delayed post-test design?

A rich body of accumulated literature has shown the positive effect of visualised intonation training on production by second or foreign language learners of English with diverse L1 backgrounds (e.g. Gorjian *et al*., 2013; Le and Brook, 2011; Hardison, 2005), but unfortunately little research included a control group trained by auditory-alone feedback, leaving the argument that visualisation of intonation is more facilitating than auditory intonation in the acquisition of L2 intonation less convincing. This brings us to the third research question of this thesis—is there any difference of the effects of auditory and audio-visual feedback on improvement of learners' comprehension ability?

We are also interested in what learners think of the training by answering a post-test questionnaire which was specifically designed for Chinese EFL learners who participated in the training procedure. Previous literature has found that Chinese university students rank intonation and other prosodic features as the most difficult aspects of pronunciation and segments the least, and they yearn for more pronunciation training (Lu, 2010). However, researchers seem to pay little attention to L2 students' own perception of their needs for pronunciation (Derwing and Rossiter, 2002). Participants' feedback and comments will be therefore insightful for evaluation of the training methods and future research.

3.2 Chinese participants

This section aims to provide a comprehensive account of the English proficiency of the Chinese participants and how they were grouped according to their proficiency levels for the experiment.

3.2.1 Demographic information

A total of 60 Mandarin Chinese learners of English (5 males, 55 females) participated in the project, all of whom had studied English more than six years, as English is a compulsory module at secondary schools (three years of junior and three years of senior) across China; some had even more than ten years of learning experience if their first exposure was at primary school and they continued to study English for a Bachelor degree. The participants were all studying English-related postgraduate programs at Newcastle University, i.e., TESOL and Applied Linguistics, Linguistics, Cross-Cultural Communication, or Interpreting and Translating (Chinese-English). The average age among these learners was 24 (SD=3), ranging from 19 to 34, and the mean length of their residence in the UK was eight months, with a minimum of one month and a maximum of three years.

Participants were from a wide variety of dialectal backgrounds, including Hebei, Shandong, Shanxi etc. where they speak Northern Mandarin based on which the phonological system of Standard Chinese (or Putonghua/Mandarin) was formed (Duanmu, 2002; Lin, 2007). Some came from Sichuan and Hubei where Southwestern

Mandarin is spoken pervasively as the local language, while others were originally from Zhejiang, Guangdong and Hunan where the local dialects are Wu, Cantonese, and Xiang respectively. 45 out of the 60 Chinese learners (75%) were self-claimed L1-Mandarin speakers, mostly because their parents came from different dialectal regions and thus Mandarin was the only shared language in their family. Alternatively, some of them moved to Northern China at a very young age and had been there until they came to the UK. Those who claimed that their L1 was the local dialect rather than Mandarin Chinese (e.g. Wu, Southwestern Mandarin, and Cantonese, etc.) also evaluated themselves as having a very good command of Mandarin (see Appendix I for details of the demographic information of Chinese participants).

3.2.2 Assessment of English proficiency

Learners' English proficiency, particularly reading (including grammar and vocabulary), was evaluated individually by the paper-and-pen version of the Quick Placement Test (2001) developed by Oxford University Press and Cambridge ESOL. This test has been designed to provide researchers with a time-saving yet reliable measurement (Geranpayeh, 2003) for a quick classification of ESL/EFL learners' English level (e.g. Hawkins and Casillas, 2008; Usó-Juan and Martínez-Flor, 2015). The test consisted of 60 multiple-choice questions that took about 30-40 minutes to finish. All the answers were recorded on an answer sheet and were manually marked by the researcher. The scores were matched to ALTE (Association of Language Testers in Europe) and CEFR (Common European Framework of Reference for Languages) levels as shown in table 3.1. Learners' best score for three categories of the IELTS (International English Language Testing System) test alongside the overall score was also collected to index their competence in English speaking, listening, reading, and overall skills.

Table 3. 1: The marking scheme of the Quick Placement Test in accordance with ALTE and CERF levels

No. of correct answers	**ALTE level**	**ALTE description**	**Council of Europe level**
0-10	0.1	Beginner	--
11-17	0.2	Breakthrough	A1
18-29	1	Elementary	A2
30-39	2	Lower intermediate	B1
40-47	3	Upper intermediate	B2
48-54	4	Advanced	C1
55-60	5	Very advanced	C2

3.2.3 Homogeneity of three Chinese groups

The way of grouping Chinese learners for the project was quasi-random in order to assure that learners with the same level of English proficiency were evenly distributed in three groups. Firstly, learners were stratified by their English proficiency; both IELTS scores and the Quick Placement scores were taken into account. When the number of a particular level was not dividable by three, learners at one level below but topped at this level (top one or two) would be moved one level up. Then learners within the same level were randomly assigned into three groups.

Following the quasi-randomization of participants into three groups, a one-way ANOVA was used to examine if there was any apparent difference in their English proficiency in terms of the Oxford Quick Placement Test. Having met the assumption of normal distribution (by Shapiro-wilk test, p value for all three groups were higher than 0.05) and the assumption of homogeneity of variance in different group (Levene's test, $F(2, 57) = 0.68$, $p = 0.51$),[1] the results from one-way ANOVA ($F(2, 57) = 0.23$, $p=0.80$) confirmed that the participants' English proficiency obtained by the Oxford Quick Placement Test was very similar across three groups, in which 95% of them in each group were at B2 level (scoring at 43~47 inclusively). Details of the descriptive analysis of their scores are in table 3.2.

Table 3. 2: Descriptive analysis of the score of the Oxford Quick Placement Test

Group	**N**	**Min**	**Max**	**Mean**	**SD**	**CI lower**	**CI upper**
Audacity	20	36	55	45.95	5.41	43.42	48.48
Praat	20	34	55	45.75	5.46	43.20	48.30
Control	20	35	52	44.90	4.66	42.72	47.08

(NB: N=number of participants; Min=minimum score; Max=maximum score; SD=standard deviation; CI upper and CI lower equal to 97.5% and 2.5% quantile respectively for 95% confidence interval)

[1] Shapiro-wilk normality test result of the Audacity group was W=0.97, p=0.67; the Praat group was W=0.98, p=0.91; the control group was W=0.94, p=0.26.

Since all participants were also required to provide their IELTS listening, speaking and reading score, it was necessary to test the differences of these scores among three groups. Before MANOVA was administered, a battery of tests was run in order to meet the assumptions that are required by MANOVA. By checking of variance-covariance matrices, it turned out that the variances of each IELTS test among the three groups were roughly equal,[2] while the covariances were slightly different for each combination of the IELTS tests among three groups.[3] But considering the equal size of each group, this difference could be ignored (Field *et al.*, 2012, p. 725). However, when exploring the multivariate normality of these test scores, the data from the Praat group seemed not normally distributed ($W=0.78$, $p<0.001$), while the other two groups had a normal distribution.[4]

Figure 3.1 shows the multivariate outliers in which case 21, 40, and 46 (i.e. participant id was 31, 50, and 56) deviated from the majority. Knowing that IELTS scores were only inspected as an add-on knowledge of learners' sub-level of English proficiency, these outliers were not excluded from the database, but gained an extra attention when examining their intonation ability. In this regard, a robust MANOVA was employed on the ranked data with Munzel and Brunner's (2000) method, implemented by Wilcox and his colleagues (2016) in R. The result uncovered that three Chinese groups had similar scores in terms of IELTS listening, speaking and reading tests, $F=0.90$, $p=0.45$.

[2] The variances for listening were 0.91, 0.84, and 1.03 corresponding to the Audacity group, the control group, and the Praat group. The variances for speaking were 0.29, 0.57, and 0.36. The variances for reading were 0.67, 0.84, and 1.02. None of the variance ratio was bigger than the threshold of 2.

[3] The covariance of speaking and listening were 0.31, 0.53, and 0.32 for the Audacity group, the control group, and the Praat group respectively. The covariances of reading and listening were 0.52, 0.61, and 0.85. The covariances of reading and speaking were 0.23, 0.37, and 1.02.

[4] The Shapiro-wilk normality test for multivariate result for the Audacity group was $W=0.96$, $p=0.45$; for the control group was $W=0.94$, $p=0.22$.

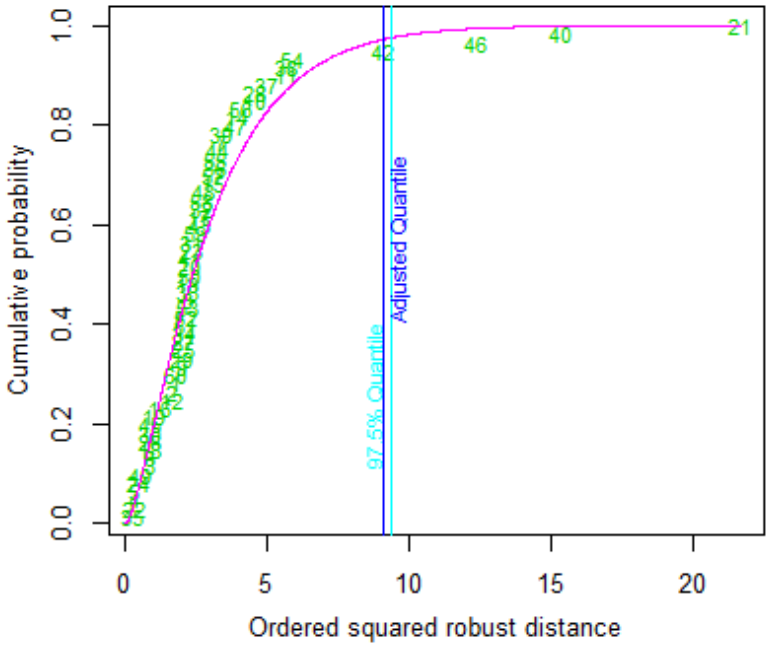


Figure 3. 1: Case numbers on the right side of the vertical line (97.5% quantile) were multivariate outliers

Table 3.3, 3.4, and 3.5 display the descriptive analyses of each IELTS test results compared separately among three groups. It seemed that Chinese learners of English were strongest at listening and reading as both skills reached an average score of more than 7 for all three groups, whereas the mean scores of speaking were all just slightly above 6. This was also evidenced by the column of CI upper and CI lower in each table which show that 95% of students in each group obtained an IELTS listening score at around 6.5~8.0, a reading score at 7.0~8.0, but a speaking score at either 6.0 or 6.5.[5]

Table 3. 3: Descriptive analysis of the IELTS listening score

Group	**N**	**Min**	**Max**	**Mean**	**SD**	**CI lower**	**CI upper**
Audacity	20	6	9	7.3	0.95	6.86	7.74
Praat	20	6	9	7.7	1.01	7.18	8.12
Control	20	6	9	7.10	0.92	6.67	7.53

[5] IELTS tests are scored at 0.5 interval, i.e. 5.0, 5.5, 6.0, 6.5, …, 8.5, 9.0.

(NB: N=number of participants; Min=minimum score; Max=maximum score; SD=standard deviation; CI upper and CI lower equal to 97.5% and 2.5% quantile respectively for 95% confidence interval)

Table 3. 4: Descriptive analysis of the IELTS speaking score

Group	N	Min	Max	Mean	SD	CI lower	CI upper
Audacity	20	5.5	7.5	6.35	0.54	6.10	6.60
Praat	20	5.5	7.5	6.45	0.60	6.17	6.73
Control	20	5.5	8	6.40	0.75	6.05	6.75

(NB: N=number of participants; Min=minimum score; Max=maximum score; SD=standard deviation; CI upper and CI lower equal to 97.5% and 2.5% quantile respectively for 95% confidence interval)

Table 3. 5: Descriptive analysis of the IELTS reading score

Group	N	Min	Max	Mean	SD	CI lower	CI upper
Audacity	20	6.5	9	7.73	0.82	7.35	8.11
Praat	20	6	9	7.55	1.01	7.08	8.02
Control	20	6	9	7.33	0.92	6.90	7.76

(NB: N=number of participants; Min=minimum score; Max=maximum score; SD=standard deviation; CI upper and CI lower equal to 97.5% and 2.5% quantile respectively for 95% confidence interval)

Based on the statistical results of the above-mentioned calculations, it was confirmed that three groups of Chinese participants were homogenous from the perspective of their overall English proficiency.

3.3 British participants

Ten self-claimed native speakers of Received Pronunciation or Southern British English (four males, six females) were recruited for this experiment. They were aged from 19 to 40 with an average of 27 (±8). They were either students or lecturers at Newcastle University and were all properly trained in Phonetics. Eight of them were monolingual speakers of English; one was also a professional speaker of French while the other spoke Japanese at a communicative level.

3.4 Intonation training materials

The instruction and practice materials including audio recordings for the present study were selected from the following two books:

> Wells, John C. (2006) *English Intonation: An Introduction*. Cambridge: Cambridge University Press.
>
> Carr, Phillip. (2013) *English Phonetics and Phonology: An Introduction*. 2nd edition. Wiley-Blackwell.

The major reason of targeting the above books was twofold. First, their descriptions of English intonation were based on the British approach which seems to be more learner-friendly from a pedagogical point of view. Second, they both supply an abundance of audio recordings from multiple native speakers to accompany every type of introduced intonation patterns in the book.

The selection process was underpinned by a communicative theme. This means that the major concern was teaching the elements that heavily account for the intelligibility and learners' communicative competence; intonational constituents that seem to be of merely attitudinal or emotional variations in functions were thus left out, e.g. the patterns of pre-head in which the difference between the high and low pre-head is associated with the degree of emphasis. Chinese learners' difficulties along phonetic, phonological, and semantic dimensions reviewed in Chapter II were given with extra emphasis in choice of materials, for instance,

lowering the pitch of the accented syllables in interrogatives, phrasing principles for restrictive and non-restrictive clauses, etc. Section 3.4.1 and 3.4.2 features the content included in the instruction and the practice materials respectively with justifications from the literature.

3.4.1 Intonation instruction material

The selected content was mindfully arranged in six sessions each of which had a focus on a particular key element of English intonation except the first and the last session. Session one was intended to raise the awareness of the importance of learning intonation to convey to teachers and students which elements of their speech are crucial for intelligibility and comprehensibility and thus to know where they should pay more attention to (Ramírez Verdugo, 2006; Cauldwell, 2013). Chinese teachers and students alike seem to urgently need this (Wang, 2013).

The session also gave a brief introduction to Halliday's (1970) 3T intonation system (tonality, tonicity and tone), along with a taste of native recordings of contrastive patterns in terms of each T in the conveyance of different semantic and pragmatic meanings of otherwise identical utterances. The overall plan for the whole instruction was explicitly told to the students at the end of this session.

Session 2 was centred on nuclear tones. Three nuclear tone patterns were chosen as the learning targets—falling, rising and falling-rising, as these patterns are most frequently and consistently used by native speakers (Cruttenden, 2014) and can denote distinctive meanings that should be grasped by learners. More intricate patterns such as high fall vs. low fall, high rise vs. low rise, and complex tones like rising-falling were deliberately omitted, because for the purpose of efficient communication, there seems no need to learn intonation patterns bearing high variability among native speakers (Levis, 1999; Grabe *et al.*, 2005).

The session started with the introduction of the phonetic realisation of three nuclear tones, with particular emphasis on the rising and the falling-rising tone in which Chinese learners were found most difficult (Ji *et al.*, 2009; He *et al.*, 2012a; Chen, 2013), and on the realisation of three tone patterns on longer domains (when nuclear syllable was in non-final position) for the reason that Chinese tone is syllable-dominated and Chinese learners are better at producing tones on accented syllables in phrase-final position with transfer from the L1 (Chen, 2008a). Intonational meanings corresponding to three nuclear tones were in this session more emphasized on those idiosyncratic uses in English, e.g. implicational fall-rises for things left unsaid, rises for declarative and echoic questions and non-final listed elements, and falls vs. fall-rises for negation scope as in "I won't eat anything". Apart from these usages that Chinese learners could hardly be aware of on their own, the distinctions between falling and rising tones were kept at a very abstract level as the traditional grammatical links of rising and questions, or falling and statements have been proved oversimplified (Cruttenden, 1981). Meanwhile, native speakers have been found to be very tolerant of tone varieties, as their nuclear tone choices are inconsistent (Grabe *et al.*, 2005).

Session 3 and 4 were focused on tonicity with a preference for the rules of nuclear accent placement, as discovered by previous studies that Chinese learners failed to locate sentence stresses on correct positions (He *et al.*, 2010; Wang *et al.*, 2010a; Zeng, 2017). Given that Mandarin Chinese and English share a similar functional use of accentuation for narrow focus (Li, 2012), contrastive/emphatic and new information which attract nuclear tone placement should not be a major problem for Chinese learners. Utterances without narrow focus, on the contrary, seem to impose great difficulty. Therefore, the essential rules of the unmarked nuclear placement (last lexical item, or LLI rule) were heavily instructed. Accentuation rules for grammatical structures that either lack Chinese equivalents or is absent of intonational treatment in Chinese were also illustrated with audio recordings. These structures normally disobey the LLI rule. The following cases were included in the instruction (nuclear syllable are in bold and underlined):

(1) event sentences: *The **phone**'s ringing.*

adverbials of time and place: *I had an unexpected **let**ter yesterday.*

final verbs and adjectives: *How's the **home**work going?*

phrasal verbs: *sit **down**, break **down**, bump **into**, look **after**, etc.*

empty content words: *things, people, etc.* (normally cannot do nucleus)

As for the phonetic realisation of nuclear accents, the concentration was placed on pitch change, coupled with the knowledge from the previous session on the realisation of nuclear tones. Mandarin Chinese, as reviewed in section 2.2.2, employs duration, intensity and pitch range/register as primary phonetic manipulations of focused words (Xu, 1999; Xu, 2004b; Lin and Li, 2011), most of which are consistently found in English too (Sluijter and van Heuven, 1996; Cruttenden, 1997). So the instruction on such strategies was kept to minimum, while emphasis was directed to the ways to de-accentuate words out of information centre. The literature has evidenced that Chinese learners fail to significantly reduce the duration of unstressed elements (Chen *et al.*, 2001; Setter, 2006), resulting in the serious consequence that their speech sounds less comprehensible due to evenly distributed information.

The importance of tonicity in communication was repeatedly brought up by the researcher during the instruction for two reasons: 1. it is crucial for information delivery in English. Native speakers' placement of nuclear accents seems to be uniform with little variation (Grabe *et al.*, 2005). 2. Chinese learners have bare knowledge of accentuation placement and the way they manipulate pitch is subject to L1 influence which for most of the time when they did not intend to stress some parts of their speech (e.g. Chen, 2008), the assignment of a contour tone leads to a perceptual impression that they were stressing it. In agreement with Jenkins (2007) and Grabe *et al.* (2005), having learners acquire the explicit knowledge of where and how to produce tonicity is considered as the first but most important step for better pronunciation competence.

Session 5 introduced the last essential element of English intonation—tonality or chunking. Wells (2006) states that chunking should not be worried about too much by learners as it functions considerably similarly across languages (p. 187). Chinese learners, however, have a great deal of difficulty in this regard by placing intonation boundaries in inappropriate positions (Chen, 2006; Zeng, 2017), even though syntactic structures largely overlap in Mandarin Chinese and English (Cao, 2012). A plausible explanation for such a phenomenon might be the resultant problematic production of accentuation as reviewed in section 2.3.5. Chinese learners are inclined to produce every single word with invariant duration and tone type (falling tone for most of the time), ended up with a seemingly syllable-timed rhythm in their English like they do in Chinese (Chen, 2008a). Some researchers argue against the value of teaching stress-timed rhythm of English (Crystal, 2003; Deterding, 2010). It is indeed not the nativelike rhythm that matters. What really matters and might detrimentally affect the intelligibility of learners' speech is the lack of sound phonetic manipulation of the information structure of their speech. They might possibly merely have no sense that they signal certain cues to an intonation boundary, just like they make a syllable protruded without intention by improperly changing their tone of voice. For instance from the researcher's personal experience, an utterance like "*I think she's a PhD student"* is often uttered by Chinese students with a nuclear accent on *think* and a significant pause after it, cutting the utterance unnecessarily into two separate intonation units, and leaving an impression that the speaker was emphasizing that it was his/her thinking rather than stating a fact. In a nutshell, the teaching of phrasing underlined the importance of not making boundary markers where no chunking was intended. So detailed instruction on the phonetic realisation of intonation boundaries was presented to the students, particularly on anacrusis, final lengthening and pitch reset which were sparsely found in learners' production (Chen, 2006). Syntactic structures that are absent of Chinese counterparts were mentioned on purpose, such as the chunking rules for the distinction between restrictive and non-restrictive clauses and for the reporting phrases both of which tend to impose difficulty on Chinese learners (Chen, 2008a).

The last session was conducted as more of a review session. It summarized the major principles of using 3Ts and their respective phonetic realisations with exceeding stress on the dimensions that Chinese learners were found most challenging. Then it combined 3Ts into more complex and longer utterances and discourses, and it also broadened the use to tone concord, varieties of nuclear falls and simple heads, all of which were just for the further interest of some advanced students. Full account of the instruction materials are in Appendix IV.

3.4.2 Intonation practice material

The practice material was designed to accompany the instruction for each session; thus six pieces of practice materials were created for use assisted by the computer tools, i.e. *Praat* and *Audacity*. The audio samples used in all materials were extracted from the two target textbooks Wells (2006) and Carr (2013). There was a total of 156 audio files with an average of 26 in each practice session, pertaining to words, phrases, sentences, dialogues, and longer narratives. The design for the tasks and activities on these materials was rigorously founded on pedagogical theories. The key features of the design included the following seven aspects:

a. The perception and production tasks were similarly weighted. Studies have shown that training on perception can lead to improvements in production, and vice versa (e.g. Catford and Pisoni, 1970; Hazan *et al.*, 2005).
b. Perception tasks started from the discrimination of phonetically or phonologically distinctive tones, and moved to contextualised words, phrases and longer utterances that can be semantically distinguished by using different tones (Trouvain *et al.*, 2016).
c. Perception of tonicity and tonality always involved with learners' metalinguistic knowledge. After marking down the locations of nuclear accents and intonation boundaries by listening to audio files, learners were required to check the answers and think about why this word was accented or why the utterance was chunked like this.

Answers to questions regarding intonational knowledge were always provided too (Celce-Murcia *et al.*, 2010; Trouvain *et al.*, 2016).

d. Production tasks started from the meaningless tone patterns on words and phrases and moved to richer contexts—sentences and dialogues. When tonicity and tonality moved in, production tasks always attempted to elicit learners' metalinguistic knowledge of where and why on nuclear placement and phrasing, so was for the tone choice in contexts. Answers were always provided for these types of tasks (Kurt *et al.*, 2014; Trouvain *et al.*, 2016).
e. A considerable amount of native input was provided (Doughty and Long, 2003).
f. The tasks and exercises were arranged hierarchically, from the focusing on simple words to phrases then to more contextualised sentences and dialogues. Intonational elements practiced in these tasks progressed from the easier ones to more difficult ones for learners, e.g. from falling tone to rising then to falling-rising; from LLI nuclear placement to more marked positions, etc. (Celce-Murcia *et al.*, 1996; Nicolaidis *et al.*, 2015).
g. Learners' difficulties were always the top concern for designing the tasks (Nicolaidis *et al.*, 2015). For instance, in activities requiring comparisons of learners' own production and the native recordings, learners were asked to pay attention to the difference between the pitch variations of nuclear accents and unstressed syllables as this was a typical issue that they were rarely aware of (Wang *et al.*, 2011a).

See Appendix V and Appendix VI for the practice materials for the Praat and Audacity group respectively.

3.5 Process of intonation training

The training process in this experiment lasted for three weeks with two sessions each week. Each session was composed of a 60-75 minutes of instruction and a 60-75 minutes of computer-assisted self-practice. The delivery of the instruction was more of a lecture during which the researcher introduced the intonational knowledge on PowerPoint based on the

materials presented in previous sections, so as to heighten Chinese learners' awareness of intonation. There were also interactions between the researcher and the students by means of question and answer to ensure that they fully understood every piece of information. The audio recordings of native samples presented during the instruction was never visually displayed in order to guarantee that the students who were about to practice on *Audacity* had no experience of seeing the intonation contours. Intonation diacritics for nuclear tones that the traditional British approach has used (see section 2.1.1) were not adopted in the instruction, nor in the practice, for not all of them were unable to unambiguously indicate where the pitch movement ends or makes turns as in ↗, ↘↗, ᷈, ᷉, ᷌, etc. which was always inserted before the nuclear syllables.

The practice following the instruction was monitored by the researcher. Participants who were pre-assigned into the *Praat* and the *Audacity* group received the same practice material which was printed out on paper. They were told to follow the instructions on the paper handout to finish every task, one step at a time, by listening to the relevant sound file that was provided to them. With the help of *Audacity*, participants could listen to the sound file as many times as they needed. Then they recorded themselves reading out the model utterances for as many times as they wished. The previously recorded files were not erased and could all be shown on the screen in a chronological order. So that the participants could replay any repetition of their own production and meanwhile play back the model recordings, either the whole or selected fragments for comparison. When they finish the first task, they moved on to the next by removing all the recordings they had saved for and opening the target sound file for the next task, and repeated the whole process.

For students who used *Praat*, all the model utterances were provided with corresponding textgrids which contained four tiers: the first tier for sentence transcription, the second for segmented words, the third for nuclear tones (or non-nuclear elements depending on the focus of that session), and the fourth with comments on any abnormal displays of the pitch curves (e.g. octave errors caused by creaky sounds) and the important cues that indexed the

realisation of the nucleus and the phrase boundaries. Participants in this group were additionally aided by visual displays of intonation when doing tasks on the paper handout. When comparison of their own production and native samples were required, they were able to open the visual displays of both. They might have needed multiple productions for a particular task if their intonation curves did not visually resemble the model.

Figure 3.2 is a screenshot of one sample utterance with accompanying textgrid.

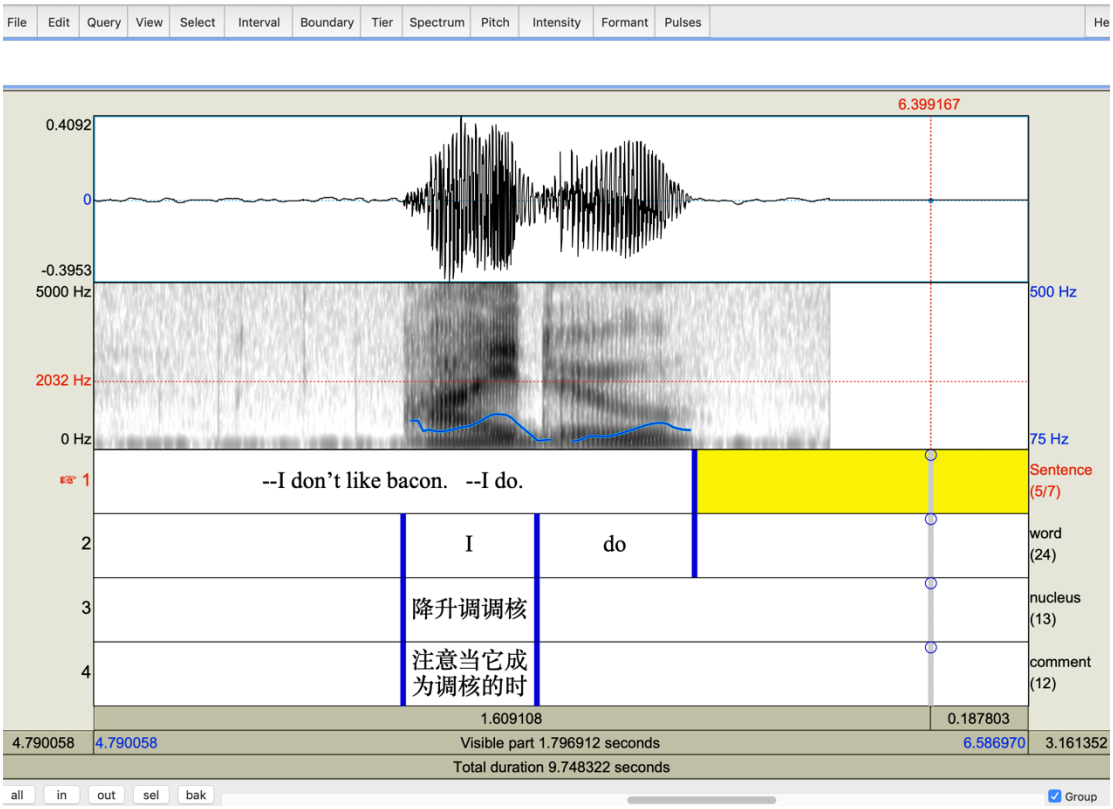


Figure 3. 2: a sample of the practice utterance the Praat group used

All the basic techniques that both groups needed for the use of *Audacity* and *Praat* were carefully guided by the instruction in the first practical session with extra time (see session 1 in Appendix V and VI for the instruction on using Praat and Audacity).

3.6 Comprehension experiment design

This section presents the experimental design for the comprehension test which was administered for the purpose of investigating Chinese EFL learners' understanding of semantic and pragmatic meanings of English intonation before, immediately after, and two months after the intonation training.

3.6.1 Sources of the stimuli

The utterances used in the comprehension tasks were carefully selected from three course books on intonation by Wells (2006), Bu (2003), and Carr (2013) respectively. The selected target utterances had to, firstly, be using as simple vocabulary as possible so that listeners would not misconceive the intonational meaning due to a limited understanding of the lexico-semantic contents (Cruz-Ferreira, 1989).[6] Secondly, the utterances had to cover a wide range of sentence structures that might cause trouble for Chinese participants to decode the contrasting meanings as previous studies have found. Thirdly, the utterance itself was literally ambiguous in terms of semantic, syntactic, or pragmatic meanings which bore the possibility of generating a pair of contrastive intonation patterns. Lastly, each feature in the 3Ts system had to be assessed by at least five utterances. In all these regards, 20 utterances (including individual sentences and two-line dialogues) were selected and numbered from 1 to 20, each of which was generated with two versions of intonation patterns delivering meaning contrasts. Eight out of 20 utterances targeted accentuation, seven phrasing and five tone; they were used for the pre-test. Due to the short period of the training sessions (three weeks) within which learners might have retained the testing effect, utterances used for the post-test were different but functionally matched to those for the pre-test. Utterances for the two-month delayed post-test were identical to the post-test.

[6] The vocabulary of testing items in all experiments was not beyond the level of B2, checked via an online platform of English vocabulary profile. The link is: http://vocabulary.englishprofile.org/staticfiles/about.html.

Examples from the target utterances (See appendix III for all the stimuli):

(1) Meaning contrasts via different phrasing:

I've washed | and ironed the clothes. (=I've got washed, and I've ironed the clothes.)

| I've washed and ironed the clothes. | (=I've done washing the clothes and ironing them.)

Meaning contrasts via different accentuation:

We're planning to fly to Italy. (In response to: Are you going to take a train to Italy?)

We're planning to fly to Italy. (In response to: Are they planning to fly to Italy?)

Meaning contrasts via different tones:

--Sophie's brought her friend along.

--Who? (Who's brought her friend along?)

–Who? (Sophie's brought who?)

3.6.2 Recordings of the stimuli

Two native speakers (one male and one female) of Received Pronunciation who had expertise in English phonetics produced the stimuli for this experiment in the soundproof booth of the speech lab at the Speech and Language Sciences Section of the School of Education, Communication and Language Sciences at Newcastle University. They were asked to read out each intonation pattern of each utterance for three times at a normal speaking rate. Their voices were recorded via a Behringer ECM8000 microphone (15-20 kHz frequency response) connected to the Edirol R-44 recorder with a sampling rate of 44 kHz, 16 bits. The best recording of each pattern of each target utterance was extracted and saved as individual sound files into a folder. The intensity of all these sound files was normalised to 75dB in Praat.

3.6.3 Implementation of the test in DMDX

DMDX was employed to implement this experiment. Each of the 20 target sentences was repeated three times, which means one pattern of each sentence was randomly chosen to be repeated twice while the other once. The reason behind repeating sentence instead of pattern for three times was simply to reduce the time of the test to a reasonable length, as the current design occupied 40 minutes already. This yielded 20×3=60 trials for the experiment. All these trials were quasi-randomised in order to avoid any utterance appearing twice in a row. The final order of the trials was identical for all listeners. Three practice trials were provided preceding the test trials, and they were not included in the test items. The pause between each trial was 2 seconds. Within each trial, the tested item alongside two meaning glosses appeared on the computer screen in black with a white background, while the target utterance was highlighted in red in order to direct listeners' attention to it. Listeners had10 seconds to familiarize themselves with the lexis. Note that the comas in the target utterances were removed, as it was found that commas cuing prosodic boundaries facilitated comprehension (Jun and Bishop, 2015). Followed by a beep sound, the audio of the stimuli was played. Listeners were asked to choose the correct meaning that they perceived by pressing the labelled key (A or B) corresponding to the meaning glosses. If they did not respond within 10 seconds, the test jumped onto the next trial. The number of the trial was also shown on the top left corner of the screen to make the listeners aware of how far they were.

Unlike some studies in which participants were allowed to listen to the stimuli multiple times and thus only a single answer was generated from each stimuli (e.g. Cruz-Ferreira, 1987; He *et al.*, 2012b), this experiment forced listeners to make a choice as soon as the stimuli was played, and as with the repetition of target sentences, three answers from each sentence were elicited. Such a design was deemed able to account for the randomness of listeners' choice by conducting robust statistical analyses, such as mixed effects models (Cunnings and Finlayson, 2015). Data from previous studies, on the contrary, could not necessarily circumvent the

possibility that a large proportion of correct answers by participants may have been due to chance, as admitted by Cruz-Ferreira (1989).

3.6.4 Procedure of the experiment

The comprehension task was delivered in a soundproof booth in a speech lab. Listeners came in individually at the time of their appointed slot. They were explicitly told that the experiment was about English intonation. If listeners were unsure about the definition of intonation, very basic knowledge was provided without mentioning any technical terms. Listeners were seated in front of a laptop, wearing noise-cancelling headphones (Bose QuietComfort 35) at a comfortable volume. Before starting the experiment, the written form of the instructions on how to do the task was displayed on the screen. The researcher made sure that listeners understood the instructions before they moved to the practice trials by providing oral explanations in Chinese. The researcher did not leave the booth until listeners finished the practice trials and demonstrated that they fully understood how to do the task. When the last trial was done, a message on the screen appeared, instructing listeners to let the researcher in and save their answers. Native listeners and Chinese participants from the control group received financial compensation, while Chinese participants from the treatment groups compensated by the potential benefits they got from the intonation training.

3.6.5 Chinese participants self-reported intonational knowledge

Chinese participants were required to fill in an online questionnaire (see Appendix 2) after they finished the pre-test. This questionnaire aimed to elicit learners' self-reported knowledge of intonation via rating nine statements about the extent to which they have mastered English intonation, alongside nine multiple choices which asked for their previous experience with learning English pronunciation back in China and their attitudes towards the teaching and

learning of pronunciation (including intonation). A total of 53 (88% response rate) online answer sheets were collected.

In general, nearly 80% of the respondents did not think they could properly manipulate phrasing, accentuation, and tone to convey meanings, but they believed in their fair perceptual ability of these features (slightly above 50% of responses). All participants acknowledged the importance of intonation in their daily communication, especially when talking with native speakers, but only one participant felt confident about his/her intonational skill.

With regard to participants' previous experience of learning intonation, few of them (less than 20%) had had systematic intonational training, although more than 90% had attended modules specifically for English pronunciation or for English in general, integrated with pronunciation teaching. This reveals the fact that the situation regarding the marginalised teaching of prosodic aspects in mainland China has not changed throughout the decades (Zhang, 2004; Lu, 2010b), no matter how advanced the theoretical research on prosody has become, and no matter how much the linguists and practitioners call for the priority of teaching and learning of prosody (Lu, 2010). The reasons of the existence of such a gap have been thoroughly discussed by Gut *et al.* (2007). The leading reason is the scarcity of empirical studies that explore the effect of teaching and learning of prosodic features on learners' perception (or comprehension) and production of these features; another is the lack of evidence of any enhancement of intelligibility and comprehensibility of learners' speech, which has subsequently resulted in the failure of translating theory into pedagogy.

Learners themselves admitted that the pronunciation training they had received did not really help improve their pronunciation (nearly 50% of the responses), but 80% of them believed that pronunciation was at least equally important as other aspects of English, e.g. grammar and vocabulary. More interestingly, when they were asked in what aspects they think they had improved by the pronunciation training, they weighted segments and suprasegments equally. However, when it came to "what aspects do you think that need to be improved", almost 90% of respondents gave weight to suprasegmentals, such as lexical stress, intonation, rhythm,

linking, etc. over consonants and vowels. This suggest that most of the Chinese learners of English felt that their suprasegmentals were worse than their segments and they were willing to learn how to understand and use suprasegmentals to increase their competence in oral English.

3.6.6 Post-test questionnaire

A post-test questionnaire was developed to elicit Chinese participants' perception of intonation knowledge and evaluation of their own skills after the training (Part 1), and this part was exactly the same as in the pre-test questionnaire, aiming to explore if there was any difference of their self-evaluation. Part 2 was made up with six multiple-choice questions that asked for their feedback and comments on the training sessions. The collected data from the post-test questionnaire is presented in Chapter V when referring to the discussion of the comprehension experiment. See Appendix VII for this questionnaire.

Chapter IV : Data analyses and results of the experiment

This chapter describes the process of data analysis in details using advanced statistical methods, and it is organised into four sections. Section 4.1 presents the results of the pre- and post-test by native speakers. Section 4.2 compares the pre-test results between the Chinese and native participants, followed by the results from Chinese participants' performance in the pre-text between different levels of English proficiency in section 4.3. Section 4.4 presents the training effects by comparing the pre-, post-, and delayed post-test between the three experimental Chinese groups.

4.1 Comparison of the pre- and post-test by native listeners

The main reason behind conducting such a comparison was to rule out the items that native listeners had failed to identify, as these items were out of interest to the current research questions. By doing so, the tests themselves were comparable across participants. So a series of generalized linear mixed effect models (GLMM) were fitted to the dataset of native listeners' performance in R (R Core Team, 2014).[1] By model comparisons using log-likelihood ratio test, the optimal model was fitted with *time* (pre- and post-test) as the main fixed effect, controlled by *condition*.[2] Random effects included by-subject and by-item random intercept and by-item random slope for *time*.[3] Any other additional fixed effect and random effect did not significantly improve the model. A total of 1200 (10×60×2) data points were counted into this model. It was found that time had a significant effect on listeners' judgement on prosodic meanings ($\chi^2(1)=3.91$, $p=0.05$), so did condition ($\chi^2(4)=10.82$, $p<0.05$), in which by-item variance ($R^2=3.78$, SD=1.95) was much larger than by-subject

[1] It models categorical outcomes predicted by both fixed and random factors. It was implemented by *lme4* package in R.
[2] The interval variable "condition" had five levels: new1, new2, new3, old2, and old3. The number refers to the occurrence of the sentence, while "new" and "old" refers to the intonation pattern. So new1 means this is the first appearance of the sentence in the comprehension task with a particular intonation pattern, and new2 means this is the second appearance of the same sentence but in the other intonation pattern. Old3 means this is the third repetition of the sentence spoken in an intonation pattern that is new.
[3] The fitted model was prepost11.11 <- glmer (judgement ~ time + condition + (1|id) + (1+time|item), data=justNative, family = "binomial", control=glmerControl (optimizer="bobyqa", optCtrl=list(maxfun=1e6)))

variance (R^2=0.15, SD=0.38). This suggests that native speakers were more or less similar in their sensitivity to intonational meanings and their sensitivity was stronger to some items than to others.

Therefore, with a closer inspection of the correct rate for each item across participants in pre- and post-test, it was found that item 16 and 19 in the pre- test, and item 9 in the post test earned a low accuracy rate (below 70%), while the rest of the items remained correct for 80%~100% of the responses. Then the same GLMM was performed again but for an updated dataset which excluded item 9, 16, and 19. This time there was no significant difference between pre- and post-test any more, as the effect of *time* was non-significant $\chi^2(1)=2.81$, $p>0.05$, and neither was the effect of *condition* $\chi^2(4)=9.23$, $p>0.05$.

4.2 Comparisons between native and Chinese participants for the pre-test

The first impression of learners' ability to comprehend intonation meanings was obtained from graphing the overall correct rate (averaged by *condition*) of the comprehension task for all four groups of participants, as shown in figure 4.1, in which item 16 and 19 were excluded. It is apparent that native listeners were more or less able to identify the correct meanings of all items, as the correct rate for each feature reached more than 90 percent. Tonal contrasts in particular were almost perfectly identified. As for Chinese listeners, regardless of which group they came from, there was 25% to 40% chance that they picked the wrong meaning for the target utterance. In addition, the performance of each intonation feature by the three Chinese groups was followed a similar pattern; namely *phrasing* seemed to be the most difficult feature for all Chinese participants to disambiguate, while accent and tone were equally better but still worse than the performance of native speakers.

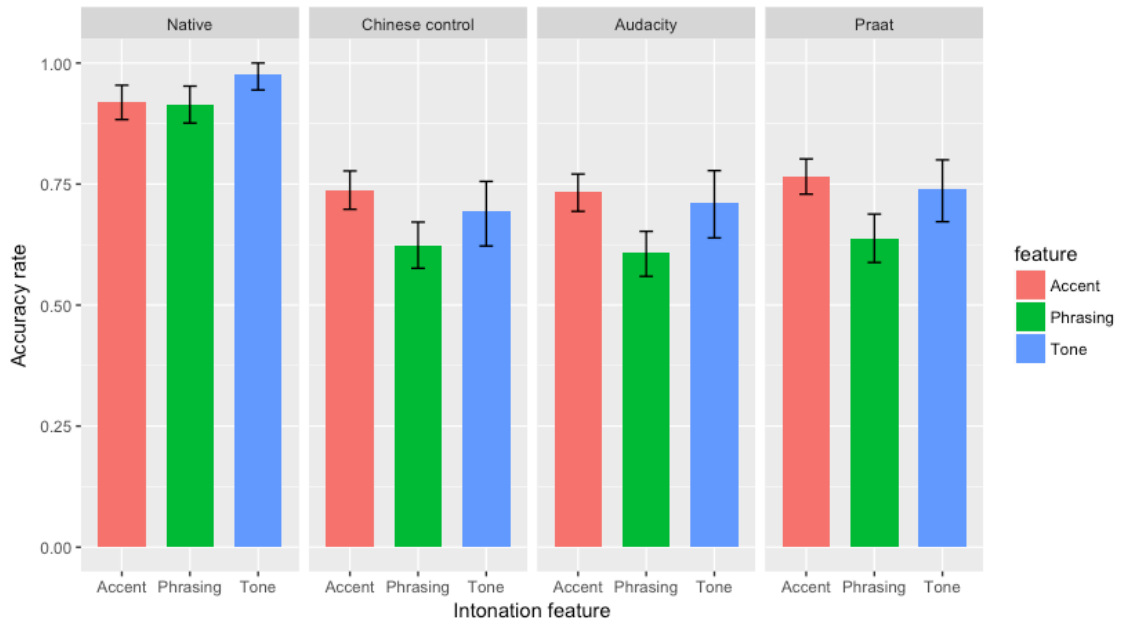


Figure 4. 1: The mean accuracy of the comprehension task in the pre-test (averaged by *condition*)

To further visualize the performance of each intonation feature in each trial condition, a line graph was created in Figure 4.2, which interprets the correct rate for the three features as a function of *condition* by all groups.

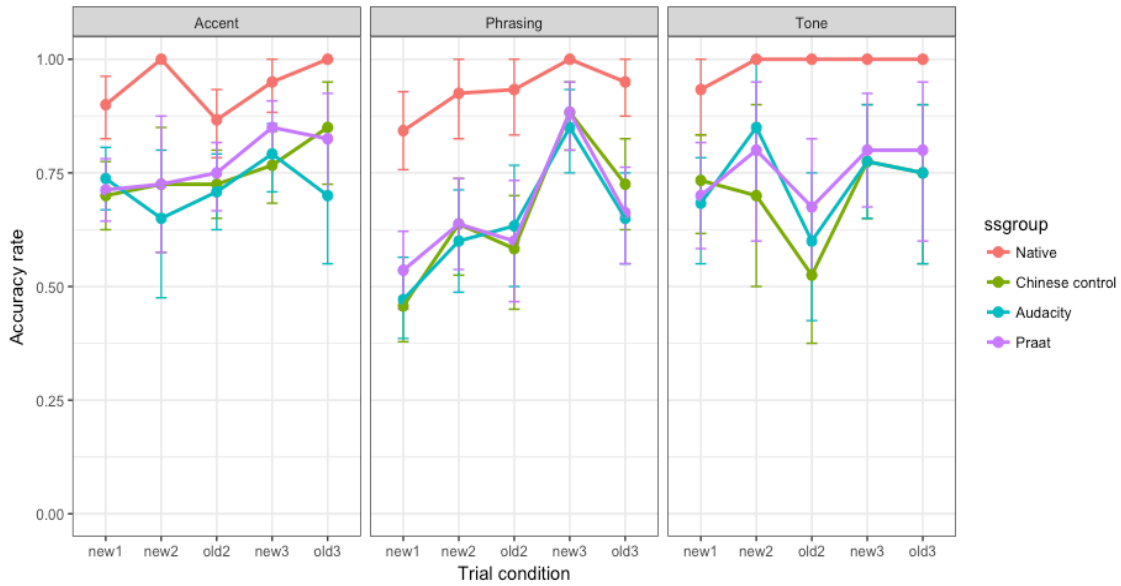


Figure 4. 2: The accuracy rate of each intonational feature as a function of condition of the comprehension task in the pre-test

Again, native listeners were better than Chinese listeners in any condition, as the correct rate of each condition for the native speakers was much higher than for the Chinese, particularly for phrasing and tone. The error bar of each condition for native listeners did not overlap with those for Chinese listeners, indicating a very likely significant difference between native and Chinese listeners in terms of every condition. It is also worth noting that not only were the accuracy rates for Chinese groups very close to each other in every condition, but also the error bars largely overlapped in each condition, which suggests that all three Chinese groups had very similar abilities. In other words, there seems to be no significant difference in the ability to disambiguate intonation meanings among Chinese groups.

Since three Chinese groups' performances on the task were similar, it was deemed clearer to explore the detailed differences between Chinese learners on the one hand, and native speakers on the other, if the three Chinese groups were treated as a whole. Thus, for the following inferential statistical analysis, the original group variable *ssgroup* was coded into a new variable called *bgroup* which had only two levels, one for the native listeners (the base level) and one for the Chinese listeners. A string of Generalized linear mixed effects models (GLMMs) were fitted using lme4 package (Bates *et al.*, 2015b) in R (R Core Team, 2014) to discover the underlying differences in the performance between them, as GLMMs are considered as the most suitable methods in analysing categorical outcomes when both fixed and random effects are in play (Manning, 2007; Baayen, 2008; Jaeger, 2008; Quené and van den Bergh, 2008; Bolker *et al.*, 2009).

The analyses started with a full model which included the fixed effects in question, i.e. *bgroup*, *feature*, *condition*, and two-way interactions between each two of them. The dependent variable was coded into binary responses (correct or wrong). Random effects were constructed according to the "maximal model" rule (Barr *et al.*, 2013), which requires random intercepts and slopes for all relevant fixed factors, namely in the current model, random intercept for the subjects and items, random slopes of *feature* and *condition* for the subject. This assumed that individual variations for the overall task as well as for each intonation

feature and for each condition were taken into account, and by-item variants were also accounted for in the adjustment of the intercept. This was deemed appropriate because listeners might have different ability to comprehend the intonation meanings in general. Meanwhile, different conditions of each intonational feature might bias their sensitivity to the meanings. Testing items themselves were also not assumed to be of identical difficulty in terms of comprehension.

However, the model failed to converge. This is often the case partly because of the overcomplicated random effects (Cunnings and Finlayson, 2015). So a series of subsequent models with progressively simplified random effects were fitted until reaching convergence. The p values generated from such simplified models would "contribute to the performance metrics for maximal models" (Barr *et al.*, 2013, p. 266), thus being considered valid. The simplification process followed the instructions by Barr *et al.* (2013) and Cunnings and Finlayson (2015). Then the random effects used in the final model were by-subject and by-item random intercept. Although the model including the random slope of *feature* for subject was also identifiable, it did not significantly improve the model fit ($\chi^2(5)=10.26$, $p=0.06$) when comparing the model with it and without.

With regard to the fixed effects, there was a possibility for there to be an effect of the three-way interaction of *bgroup*, *feature* and *condition* via the visualisation of the distribution of accuracy rate for each condition of each feature across all groups as shown in Figure 3. Therefore, a model added with this fixed effect was attempted but ended up with a failure of convergence, plausibly due to the limited size of the dataset for such a complex construction (Bolker *et al.*, 2009). As a result, and by keeping the random structure intact, models with gradually reduced fixed effects were created. Model comparisons were done by the anova () function in R, which uses likelihood ratio test to opt for the optimal model which fits the data

most. As any reduction of the effect failed to fit the data better, the fixed effects in the final model reverted to the original proposal. The full model was:[4]

```
pre3.3.b <- glmer (judgement ~ bgroup + feature + condition + condition:bgroup +
condition:feature + bgroup:feature +  (1 |id) + (1|item), data=justPre.up, family =
"binomial", control=glmerControl (optimizer="bobyqa", optCtrl = list (maxfun=1e5)))
```

The results demonstrated that native listeners were in general significantly better than Chinese listeners in identifying intonation meanings (b=-1.28, SE=0.34, z=-3.73, p<0.001). In order to obtain the global effect of *bgroup*, *feature*, *condition*, and their interactions, a mixed () function in afex package (Singmann *et al.*, 2016) was used to get the Wald chi-square and p value for each main effect.[5] It was found that *bgroup* had a main effect on the overall judgement of the intonation meaning ($\chi^2(1)=87.10$, p<0.0001). Although *feature* had no main effect ($\chi^2(2)=0.40$, p=0.82), its interaction with *bgroup* did ($\chi^2(2)=8.45$, p=0.01), and so did its interaction with *condition*, $\chi^2(8)=48.64$, p<0.0001; this suggests that native and Chinese participants' discrimination of intonation meaning varied across different intonation features, and listeners' comprehension of each intonation feature was affected by whether it was the 1st, 2nd or 3rd appearance of a particular item as well as whether it appeared as a new or old intonation pattern of that item. There was also an overall effect of *condition* ($\chi^2(4)=26.14$, p<0.0001), but no effect was found for its interaction with *bgroup* ($\chi^2(4)=2.96$, p=0.56). This indicated that *condition* did elicit varying sensibility to intonational meanings, but its effect seemed constant across listeners no matter whether their L1 was English or Chinese.

Since the model cannot compute the three-way interaction of *bgroup*, *feature* and *condition*, it was assumed that no effect was found for such interaction. Thus the main effect of the two-way interaction of *bgroup* and *feature* allowed us to compare the differences between groups in terms of each intonation feature, since the effect of feature on both native and Chinese listeners was not constrained by *condition*. To do so, nine pairs of planned contrasts were

[4] No outlier was detected by visually plotting fitted values against the residuals generated from the model.
[5] The command used for global effects was: pre3.3.b.afex <- mixed (judgement ~ bgroup + feature + condition + condition:bgroup + condition:feature + bgroup:feature + (1 |id) + (1|item), data=justPre.up, family = "binomial", control=glmerControl (optimizer="bobyqa", optCtrl=list(maxfun=1e6)), method = 'LRT').

administered by lsmeans () and rbind () functions in lsmeans package (Lenth, 2016) in R,[6] which compares the Least-squares means predicted by the fitted GLMMs. P values were adjusted using mvt (multivariate *t*) method for 9 tests. The set-up of these planned contrasts was necessary because the research of interest not only lay between native and Chinese participants, but also within the two groups regarding each intonation feature, such as Chinese phrasing vs. Chinese accent, native accent vs. native tone, etc. Traditional post hoc pairwise comparisons were redundant as some pairs were not necessary, e.g. native tone vs. Chinese phrasing, which might distort the adjusted p value for each pair. The details of the results are shown in table 4.1.

Table 4. 1: comparisons between and within native and Chinese listeners on different features

Contrasts	Estimate	SE	z	p value	Sig.
Native accent--Chinese accent	-1.67	0.31	-5.31	<.001	***
Native phrasing--Chinese phrasing	-2.58	0.34	-7.50	<.001	***
Native tone--Chinese tone	-3.28	0.76	-4.34	<.001	***
Native accent--Native phrasing	0.65	0.67	0.98	0.89	N/A
Native accent--Native tone	1.30	1.04	1.25	0.74	N/A
Native phrasing--Native tone	0.65	1.06	1.61	0.98	N/A
Chinese accent--Chinese phrasing	-0.26	0.55	-0.47	0.99	N/A
Chinese accent--Chinese tone	-0.31	0.71	-0.44	1.00	N/A
Chinese phrasing--Chinese tone	-0.05	0.73	-0.07	1.00	N/A

[6] The commands of setting these contrasts were: b.f <- pairs (lsmeans (pre3.3.b, ~ bgroup | feature)), f.b <- pairs (lsmeans (pre, ~ feature | bgroup)), rbind (b.f, f.b)

(NB: results were averaged over *condition* on the log odds ratio scale. P values were adjusted by mvt (multivariate t) method.)

A clear significant difference between native and Chinese participants in terms of the performance of each intonation feature can be gleaned from table 5.1, whereas neither group possessed different ability to decipher the meaning encoded by accentuation, phrasing and tone respectively. It was not surprising that the predicted mean accuracy rate for all intonation features by native listeners was high, up to more than 90%, particularly for phrasing (97.1%±1.5% SE) and tone (98.5%±1.4% SE). Although their sensitivity to accent was lower than for phrasing and tone (94.6%±2.4% SE), both differences proved to be non-significant (between accent and phrasing: ß=0.65, p=0.89, between accent and tone: ß=1.3, p=0.74), and there was no difference between phrasing and tone (ß=0.65, p=0.98). The same results were found for Chinese listeners. The observed correct rate for phrasing by this group was the lowest compared to accent and tone from Figure 5.1, but the predicted correct rate for it (71.8%±8.1% SE) was almost equal to that for tone (70.7%±12.6% SE), but this difference was not significant, and neither were any other comparisons within Chinese listeners: the difference between phrasing and accent was ß=-0.26, p=0.99, and the difference between accent and tone was ß=-0.31, p=1.00, even though the predicted mean accuracy for accent seemed higher than phrasing and tone (76.7%±6.7% SE).

In short, native listeners were able to discriminate intonational meanings cued by each of the features, and the variations among them were nearly unidentifiable due to a ceiling effect. The predicted probability for the natives in this study was likely to represent the whole population whose capability of comprehending intonation meanings should be perfect, as the range of the 95% confidence intervals of each feature was noticeably narrow. Chinese learners of English, however, while predicted to be able to decode intonational meanings above chance, still fell behind native speakers in this aspect of linguistic ability, as the odds of responding correctly to accentual contrasts by a native listener was 5 times higher than that by a Chinese listener,

13 times higher for a phrasing-carried meaning, and 26 times higher for a tone-bearing meaning. More importantly, the variations among Chinese listeners was tangible for all three features, tonal contrasts in particular, as the lower boundary of 95% confidence interval was even below chance. This implies that the power of prediction for this group was less accurate than that for native speakers (see Table 4.2 for the full details of the predicted probability of the comprehension task by both groups). An explanation proposed was the Chinese students who participated within each group were of diverse levels of English proficiency; the different performance between advanced learners and less advanced ones, if there was any, might had been even out as a result of grouping. This will be delved deeper in later sections.

Table 4. 2: Predicted probability of native and Chinese participants' comprehension of intonation features

Feature	Group	Predicted probability	SE	asymp.LCL	asymp.UCL
Accent	**Native**	0.946	0.024	0.873	0.978
	Chinese	0.767	0.067	0.612	0.873
Phrasing	**Native**	0.971	0.015	0.924	0.989
	Chinese	0.718	0.081	0.537	0.848
Tone	**Native**	0.985	0.014	0.908	0.998
	Chinese	0.707	0.126	0.423	0.888

(NB: results were averaged over the levels of condition. Confidence level used 0.95. Intervals were back-transformed from the logit scale)

Since *feature* and *condition* interacted and exerted an effect on the identification of sentence meanings, a sequence of line graphs was made to illustrate the difference between each combined situation (Figure 4.3). Participants from the native and Chinese group reacted to

each condition in the same way, since the three-way interaction of *bgroup*, *feature* and *condition* was not found. This can be gleaned from the similar fluctuation of the lines of the two groups for each feature in figure 4.3. Comparisons were therefore done among conditions under each feature as well as between native and Chinese listeners within each combined situation.

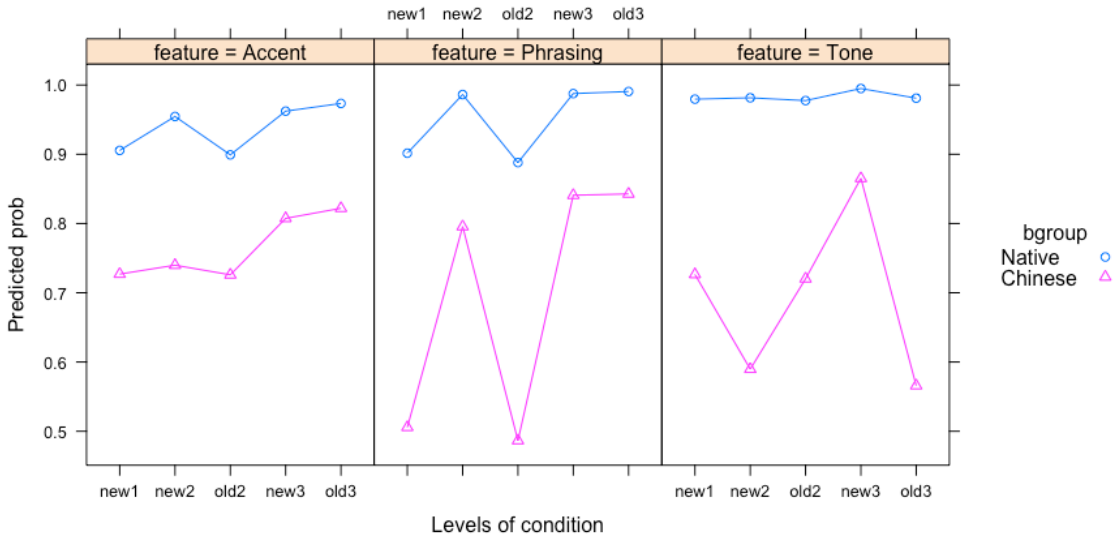


Figure 4. 3: Predicted probability of each condition between native and Chinese participants

Table 4.3 shows that for accentuation, no significant difference between any two conditions was found, although from Figure 4.3, there seemed to be some fluctuation in the accuracy rate from new1 to old3. These differences were predicted to be too subtle for any condition to make its difficulty of being comprehended stand out from the rest conditions. In other words, regardless of the L1 of the listeners, intonational contrasts embedded with different accentuations remained at a certain level of comprehensibility through re-occurrence of a particular utterance and through the alternation of new and old pattern (location of the accent).

Table 4. 3: Comparison among condition under the feature of accentuation

Contrast	Estimate	SE	z ratio	p value	Sig.
new1 - new2	0.42	0.40	1.04	1.00	N/A
new1 - old3	0.94	0.47	2.03	0.26	N/A
new2 - old3	0.51	0.53	0.97	1.00	N/A
new1 - old2	-0.04	0.25	-0.16	1.00	N/A
new1 - new3	0.72	0.35	2.05	0.24	N/A
old2 - new3	0.75	0.36	2.10	0.21	N/A

(NB: results were averaged over the levels of *bgroup*. P values were adjusted by Bonferroni method for 6 tests. Tests were performed on the log odds ratio scale.)

The patterns are more complicated for phrasing, as shown by table 5.4. Four pairs of comparisons were found to be significant with very small p values, while the remaining two pairs were not. Specifically, new2 was significantly easier to be judged than new1 (ß=1.7, p<0.0001), and so was old3 (ß=2.05, p<0.0001), but the difference between new2 and old3 was not significant (ß=0.35, p=1.00). However, old2 was not significantly easier than new1 (ß=-0.11, p=1.00), but new3 was (ß=1.90, p=0.0001), and the difference between old2 and new3 was also significant (ß=2.01, p<0.0001). This indicated that when listeners were exposed to a phrasing pattern, it did matter if they had or had not encountered this pattern before. If they had, the chance of comprehending this pattern correctly was as low as the first time they encountered it, but if they had not, meaning that the pattern was new and contrastive to the previous one, they would immediately recognise that their choice was wrong for the first pattern, and they were able to comprehend the upcoming pattern no matter if it was new or old.

Table 4. 4: Comparisons among condition under the feature of phrasing

Contrast	Estimate	SE	z ratio	p value	Sig.
new1 - new2	1.70	0.37	4.61	<0.0001	***
new1 - old3	2.05	0.42	4.87	<0.0001	***
new2 - old3	0.35	0.49	0.72	1.00	N/A
new1 - old2	-0.11	0.30	-0.36	1.00	N/A
new1 - new3	1.90	0.43	4.43	0.0001	***
old2 - new3	2.01	0.44	4.61	<0.0001	***

(NB: results were averaged over the levels of *bgroup*. P values were adjusted by Bonferroni method for 6 tests. Tests were performed on the log odds ratio scale.)

For tone, as shown in Table 4.5, the differences between each two conditions were largely insignificant, except between new1 and new3 (ß=1.14, p=0.06) and between old2 and new3 (ß=1.21, p=0.04). The former was of a marginal significance while the latter was significant at 0.05 level. This, however, did not warrant a conclusion that pattern contributed to the comprehensibility of tonal contrasts, because old3 did not significantly trigger higher accuracy rate compared to new1 and new2, and in fact, they were almost at the same level of difficulty (the difference between new1 and old3 was ß=-0.32, p=1.00; the difference between old3 and new2 was ß=-0.06, p=1.00). Instead, it is safe to conclude that the effect of condition on tonal contrasts was not as significant as it was on phrasing, indicating that listeners' judgement on meanings cued by nuclear tones seemed to remain consistent across different conditions.

Table 4. 5: comparisons among condition under the feature of tone

Contrast	Estimate	SE	z ratio	p value	Sig.
new1 - new2	-0.25	0.58	-0.44	1.00	N/A
new1 - old3	-0.32	0.61	-0.52	1.00	N/A
new2 - old3	-0.06	0.64	-0.10	1.00	N/A
new1 - old2	-0.06	0.34	-0.19	1.00	N/A
new1 - new3	1.14	0.44	2.60	0.06	.
old2 - new3	1.21	0.45	2.69	0.04	*

(NB: results were averaged over the levels of *bgroup*. P values were adjusted by Bonferroni method for 6 tests. Tests were performed on the log odds ratio scale.)

To sum up, the significant effect of *condition* that was found was only influential for *phrasing*, while accentuation and tone were hardly affected. In addition, this effect was identical for native and Chinese participants.

By comparing the Least-squares means of the accuracy rate of each combined situation between Chinese and native listeners, it was not surprisingly found that the Chinese group was significantly less capable of identifying a correct meaning in all situations than native listeners. Table 4.6 summarises the results. Some conditions tended to be very hard for Chinese listeners, such as new1 of phrasing, whose predicted mean accuracy was just at chance, 51%±10% SE, and the same applied for new2 of tone, 59%±18% SE and for old2 of phrasing, 49%±11% SE. Even though the remaining contexts were seemingly judged above chance with a mean accuracy rate about 70%, the ranges of 95% confidence intervals of these situations were notably large, which warns that these numbers might not be representative of the true population. This mirrors the results that were discussed before, that is the predicted means for Chinese participants need to be more carefully disentangled when their English proficiency is also considered. Native listerners' performance for every situation was above 90%, with only one exception for old2 of phrasing (89%). Tonal contrasts in particular was

almost impeccable, as the predicted mean accuracy for every condition under tone was above 98%. Phrasing, as revealed before, interacted with *condition*; only for new1 and old2, the predicted correct rate was slightly down to 90% and 89% respectively, while new2, new3 and old3 were all up to 99%, and these improvements were all significant. A complete summary of the predicted mean accuracy rate for each combined situation is presented in table 4.7.

Table 4. 6: Pairwise comparisons between native and Chinese in each combined situation of feature and condition

Condition	Feature	Contrast	Estimate	SE	Z ratio	p value	Sig.
new1	**Accent**	Native--Chinese	1.28	0.34	3.47	<0.001	***
	Phrasing	Native--Chinese	2.19	0.34	6.39	<0.001	***
	Tone	Native--Chinese	2.89	0.77	3.78	<0.001	***
new2	**Accent**	Native--Chinese	2.00	0.66	3.02	<0.01	**
	Phrasing	Native--Chinese	2.91	0.34	4.58	<0.001	***
	Tone	Native--Chinese	3.61	0.96	3.76	<0.001	***
new3	**Accent**	Native--Chinese	1.80	0.61	2.96	<0.01	**
	Phrasing	Native--Chinese	2.71	0.70	3.88	<0.001	***
	Tone	Native--Chinese	3.42	0.93	3.66	<0.001	***
old2	**Accent**	Native--Chinese	1.21	0.39	3.13	<0.01	**
	Phrasing	Native--Chinese	2.12	0.46	4.61	<0.001	***
	Tone	Native--Chinese	2.82	0.77	3.65	<0.001	***
old3	**Accent**	Native--Chinese	2.07	0.79	2.64	<0.01	**
	Phrasing	Native--Chinese	2.98	0.75	3.95	<0.001	***
	Tone	Native--Chinese	3.68	1.04	3.55	<0.001	***

(NB: results were given on the log odds ratio scale.)

Table 4. 7: Predicted mean accuracy of each condition under each feature by native and Chinese listeners (probabilities below 0.6 were highlighted)

Condition	Feature	Group	Predicted probability	SE	asymp.LCL	asymp.UCL
new1	**Accent**	Native	0.91	0.04	0.78	0.96
		Chinese	0.73	0.08	0.56	0.85
	Phrasing	Native	0.90	0.06	0.77	0.96
		Chinese	0.51	0.10	0.31	0.70
	Tone	Native	0.98	0.02	0.88	1.00
		Chinese	0.73	0.12	0.44	0.90
new2	**Accent**	Native	0.95	0.03	0.82	0.99
		Chinese	0.74	0.09	0.54	0.87
	Phrasing	Native	0.99	0.01	0.94	1.00
		Chinese	0.80	0.07	0.62	0.90
	Tone	Native	0.98	0.02	0.84	1.00
		Chinese	0.59	0.18	0.26	0.86
old2	**Accent**	Native	0.90	0.05	0.76	0.96
		Chinese	0.73	0.08	0.55	0.85
	Phrasing	Native	0.89	0.06	0.70	0.96
		Chinese	0.49	0.11	0.28	0.69
	Tone	Native	0.98	0.02	0.86	1.00
		Chinese	0.72	0.13	0.42	0.90
new3	**Accent**	Native	0.96	0.03	0.87	0.99
		Chinese	0.81	0.06	0.66	0.90
	Phrasing	Native	0.99	0.01	0.94	1.00
		Chinese	0.84	0.06	0.68	0.93
	Tone	Native	0.99	0.01	0.95	1.00
		Chinese	0.87	0.08	0.64	0.96
old3	**Accent**	Native	0.97	0.03	0.87	1.00

		Chinese	0.82	0.07	0.65	0.92
	Phrasing	Native	0.99	0.01	0.95	1.00
		Chinese	0.84	0.06	0.69	0.92
	Tone	Native	0.98	0.02	0.82	1.00
		Chinese	0.57	0.18	0.24	0.84

4.3 Comparisons between Chinese participants with advanced and less advanced English proficiency

In this section, the effect of learners' English proficiency on their intonational performance will be examined in order pinpoint the potential source of the individual differences that were revealed in the previous sections. The same dataset was used, excluding the native group. The target factors in the GLMMs were levelled scores of the Oxford quick placement test (with four levels, B1, B2, C1, and C2),[7] feature, condition, the two-way interaction of level and feature, feature and condition, and level and condition, and the three-way interaction of level, feature and condition. Length of residency in the UK (centred) and age (centred) were treated as control variables.[8] Attempted random effects included by-item and by-subject random intercepts, and by-subject random slopes for all fixed effects, but all random slopes by subject were redundant as they were either perfectly correlated with the random intercept for subject or they failed to significantly improved the model fit.[9] Model comparisons showed that an inclusion of each one of the following fixed effects did not significantly improve the model fit: length of the residency in the UK, learners' age, the two-way interaction of level and

[7] It was found that Chinese participant's level of this test was highly correlated with their IELTS listening, speaking, reading, and overall score. In order to avoid multicollinearity in the models, only this measure was used as it was thought to be more accurately reflecting their English proficiency at the time of testing.
[8] Centring continuous variables was necessary to avoid collinearity in model fitting. It was done by subtracting the mean from each data point.
[9] Models with the random slopes of the two-way interaction failed to converge.

condition. The inclusion of the three-way interaction led to a failure of convergence. As a result, the final model fitted for the analysis of the effect of learner's English proficiency on their task performance was:

```
prof.6.up.6 <- glmer (judgement ~ level + feature + condition + level:feature +
feature:condition + (1|id) + (1|item), data = CH.pre, family = "binomial",
control=glmerControl (optimizer="bobyqa", optCtrl=list(maxfun=1e6)))
```

The main effect of each fixed factor was obtained by afex() package (Singmann *et al.*, 2016) in R. The results are represented in table 4.8. It was apparent that *level* had a significant main effect on learners' comprehension skill ($\chi^2(3)=9.44$, $p<0.05$), and this effect was conditioned by feature, as the effect of the two-way interaction between them was significant ($\chi^2(6)=23.5$, $p<0.001$). It implied that learners' comprehension of intonational features differed according to their level of English proficiency. *Condition* alone also had a significant effect ($\chi^2(4)=65.45$, $p<0.0001$), but this effect was influenced by *feature* ($\chi^2(8)=55.1$, $p<0.0001$); this implies that learners, in general, tended to be sensitive to certain conditions of certain features. Whether advanced and less advanced Chinese learners differed in the way of deciphering intonational meanings across different features and which combined situations of *feature* and *condition* elicited higher correct rate will be explored in the following paragraphs.

Table 4. 8: Effects of all fixed factors in GLMM for Chinese participants' performance on pre-test

Effect	df	Chi-square	P value	Sig.
level	3	9.44	0.02	*
feature	2	0.30	0.86	N/A
condition	4	65.45	<0.0001	***
level:feature	6	23.50	0.0006	***
feature:condition	8	55.10	<0.0001	***

Since *feature* and *condition* interacted in accounting for learners' overall performance, customized comparisons were done between conditions under each feature by lsmeans () package in R (Lenth, 2016). Not surprisingly, *condition* had no effect for accentuation, as none of the pairs was significantly different. The strongest effect of *condition* was embodied by phrasing, as certain conditions of phrasing significantly differed from others with prominent effect size. This demonstrates that the appearance of a new phrasing pattern of certain utterances facilitated comprehension of this utterance, as the correct rate of new1 was significantly lower than new2 (ß=1.54, $p<0.0001$) and new3 (ß=1.86, $p<0.0001$), whereas old2 was not significantly better than new1 (ß=-0.21, $p=1.00$), but new3 was definitely better than old2 (ß=1.74, $p<0.0001$). Tonal pairs were found to be performed differently only between new1 and new3 (ß=0.87, $p<0.05$), and between old2 and new3 (ß=0.94, $p=0.01$), so pattern was not found to contribute to the comprehension of tonal contrasts because new2 was not judged with significantly higher accuracy than new1 (ß=-0.77, $p=0.74$). This finding was consistent with what had been found in the process of comparisons between native and Chinese performance, where new/old phrasing pattern did make a difference in terms of elicitation of a correct answer of intonational meanings from listeners, while accentual and tonal patterns did not. Table 4.9 summarizes the results of all comparisons.

Table 4. 9: Comparisons of conditions under each feature (averaged by level of proficiency)

Feature	Contrast	Estimate	SE	Z ratio	P value	Sig.
Accent	new1 - new2	0.06	0.28	0.24	1.00	N/A
	new1 - old3	0.56	0.30	1.90	0.34	N/A
	new2 - old3	0.50	0.30	1.64	0.60	N/A
	new1 - old2	0.01	0.17	0.03	1.00	N/A
	new1 - new3	0.45	0.18	2.49	0.08	N/A

	old2 - new3	0.45	0.18	2.45	0.09	N/A
Phrasing	new1 - new2	1.54	0.24	6.46	<0.0001	***
	new1 - old3	1.86	0.24	7.66	<0.0001	***
	new2 - old3	0.32	0.22	1.45	0.89	N/A
	new1 - old2	-0.21	0.24	-0.84	1.00	N/A
	new1 - new3	1.54	0.29	5.30	<0.0001	***
	old2 - new3	1.74	0.29	5.98	<0.0001	***
Tone	new1 - new2	-0.77	0.50	-1.54	0.74	N/A
	new1 - old3	-0.87	0.50	-1.75	0.48	N/A
	new2 - old3	-0.10	0.44	-0.22	1.00	N/A
	new1 - old2	-0.07	0.27	-0.26	1.00	N/A
	new1 - new3	0.87	0.30	2.92	0.02	*
	old2 - new3	0.94	0.30	3.15	0.01	**

(NB: results were given on the log odds ratio scale. P values were adjusted by Bonferroni method for 6 tests)

According to the main GLMM model fitted for the comparisons of comprehension between different levels of English proficiency, the effect of level of English proficiency was conditioned by *feature* alone, which means that learners' performance on the comprehension task varied only across intonational features, and this variation can be explained by their English proficiency. Figure 4.4 plotted the observed mean accuracy of each level of Chinese participants as a function of intonation feature. Differences can be seen from phrasing and tone, but only phrasing covered an error bar which did not overlap with others at all, i.e. C1 (the purple line). This suggests that only those learners with the highest English proficiency stood out at performing phrasing contrasts. Tonal contrasts seemed to bear a great deal of individual variation among learners (with wider error bars), but not a single group deviated from others with observable isolated error bars. Inferential analyses helped to confirm this conjecture.

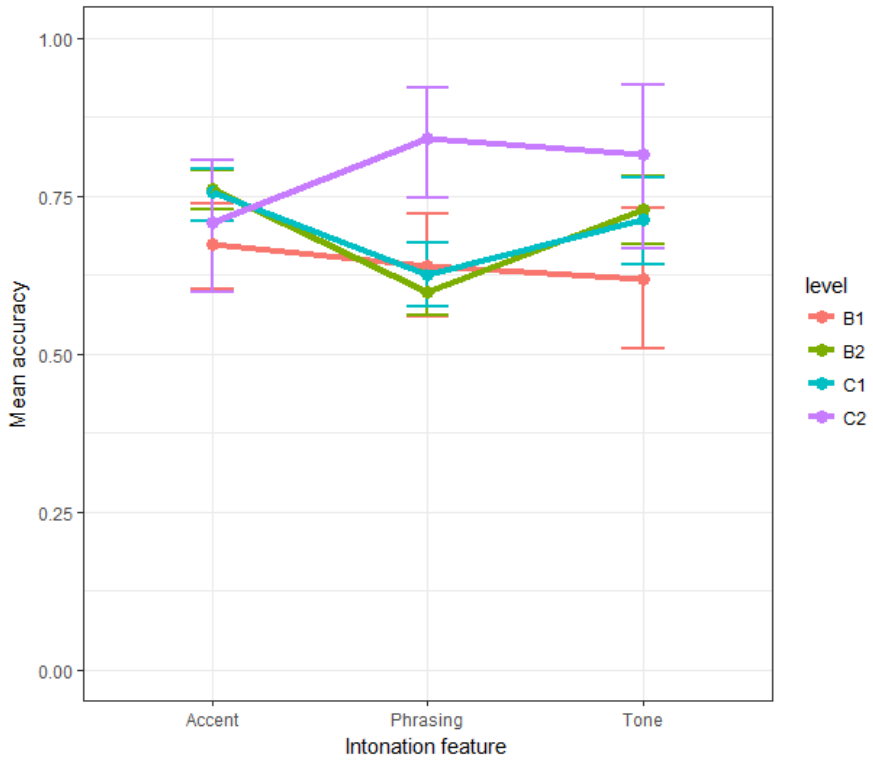


Figure 4. 4: Observed mean accuracy for different levels of English proficiency

From Table 4.10, it is clear that the effect of English proficiency was only reflected by *phrasing*. Learners with a 'very advanced' level of English proficiency (C2) performed significantly better than learners at other levels, as the difference between B1 and C2 was ß=-1.49, p=0.01; B2 and C2 was ß=-1.75, p<0.001, C1 and C2 was ß=-1.58, p<0.01. The difference between the lowest level (B1-lower intermediate) and the top level (C2) was the most significant, but learners at upper intermediate (B2) and advanced level (C1) were not significantly better than B1 level on any intonation feature.

Table 4. 10: Comparisons of predicted mean accuracy rate between different levels of English proficiency (averaged by condition)

Feature	Contrast	Estimate	SE	Z ratio	P value	Sig.
Accent	B1 - B2	-0.46	0.21	-2.22	0.12	N/A
	B1 - C1	-0.43	0.23	-1.91	0.22	N/A

	B1 - C2	-0.18	0.34	-0.52	0.96	N/A
	B2 - C1	0.03	0.16	0.20	1.00	N/A
	B2 - C2	0.28	0.30	0.93	0.79	N/A
	C1 - C2	0.25	0.32	0.79	0.86	N/A
Phrasing	B1 - B2	0.26	0.24	1.08	0.70	N/A
	B1 - C1	0.09	0.26	0.35	0.98	N/A
	B1 - C2	-1.49	0.46	-3.21	0.01	**
	B2 - C1	-0.17	0.17	-0.97	0.76	N/A
	B2 - C2	-1.75	0.42	-4.15	0.0002	***
	C1 - C2	-1.58	0.43	-3.65	0.002	**
Tone	B1 - B2	-0.55	0.32	-1.74	0.30	N/A
	B1 - C1	-0.47	0.34	-1.37	0.52	N/A
	B1 - C2	-1.10	0.60	-1.84	0.26	N/A
	B2 - C1	0.08	0.24	0.34	0.99	N/A
	B2 - C2	-0.55	0.55	-1.00	0.75	N/A
	C1 - C2	-0.63	0.56	-1.12	0.68	N/A

(NB: results were given on the log odds ratio scale. P values were adjusted by Tukey method for comparing a family of 4 estimates)

Pairwise comparisons of features within each level of participants also showed that learners at higher levels of English proficiency, just like those at lower levels, did not perform differently on different intonation features. See the complete summary of the results in Table 4.11.

Table 4. 11: Comparisons of intonation features within each level of English proficiency (averaged by condition)

Level	Contrast	Estimate	SE	Z ratio	P value	Sig.
B1	Accent - Phrasing	-0.21	0.63	-0.34	0.94	N/A

	Accent - Tone	0.44	0.82	0.54	0.85	N/A
	Phrasing - Tone	0.65	0.84	0.78	0.72	N/A
B2	Accent - Phrasing	0.51	0.59	0.86	0.67	N/A
	Accent - Tone	0.35	0.77	0.45	0.89	N/A
	Phrasing - Tone	-0.16	0.78	-0.20	0.98	N/A
C1	Accent - Phrasing	0.31	0.60	0.51	0.87	N/A
	Accent - Tone	0.40	0.78	0.51	0.87	N/A
	Phrasing - Tone	0.09	0.80	0.11	0.99	N/A
C2	Accent - Phrasing	-1.53	0.75	-2.04	0.10	N/A
	Accent - Tone	-0.48	0.95	-0.51	0.87	N/A
	Phrasing - Tone	1.04	1.00	1.04	0.55	N/A

(NB: results were given on the log odds ratio scale. P values were adjusted by Tukey method for comparing a family of 3 estimates)

4.4 Analysis of the training effects

This section focuseson how Chinese groups differed with the effect of *time* (pre, post, delayed post), controlled by *feature* and *condition*. So the fixed effects included in the major generalized linear mixed effect model were *ssgroup* (Chinese control, Audacity and Praat), *time*, the interaction of *ssgroup* and *time*, *feature*, *condition*, and the interaction of *feature* and *condition*. Attempted random effects encompassed random intercept by subject and item, random slopes of *feature*, *condition* and *time* for subject, and random slopes of *time* for item. Log likelihood ratio tests evidenced that the inclusion of the random intercepts for subject and item, and the random slopes of *feature* for subject was the best structure. This was also double checked by performing PCA (Principal Component analysis) with rePCA () function in RePsychLing package (Baayen *et al.*, 2015), which showed a reasonable distribution of variants supported by each random component, suggesting that the random structure was not overfitted (Bates *et al.*, 2015a).

The main effect of each fixed factor was gained by mixed () function in afex package (Singmann *et al.*, 2016). The results showed that *time* had a significant effect on the overall performance across groups ($\chi^2(2)=265.33$, $p<0.0001$), so did its interaction with *ssgroup* ($\chi^2(4)=161.89$, $p<0.0001$), which indicates that the effect of time was different for each group. The significant effect was also found for ssgroup alone ($\chi^2(2)=54.45$, $p<0.0001$). *Condition* had a general effect across *time* and *ssgroup* ($\chi^2(4)=69.52$, $p<0.0001$), but it was conditioned by *feature*, as the effect of the interaction of *feature* and *condition* was significant too ($\chi^2(8)=39.15$, $p<0.0001$). Since models with the addition of three-way interactions or four-way interaction all failed to be constructed, the difference between groups as a function of *time* will be explored within each feature.

A series of subsequent GLMMs were therefore run separately on a subset of the data, namely the data for accentuation, phrasing and tone. For accentuation, the model fitted *time*, *ssgroup*, *condition*, and the interaction of *time* and *ssgroup* as the fixed effects, by-subject and by-item random intercept, and by-item random slopes of *time* as the random effects. Three-way interaction of *time*, *ssgroup* and *condition* had no effect on the judgement and did not improve the model fit ($\chi^2(16)=10.5$, $p=0.39$), neither did the two-way interaction of *ssgroup* and *condition* ($\chi^2(8)=3.2$, $p=0.92$), *time* and *condition* ($\chi^2(8)=8.55$, $p=0.38$). They were, as a result, removed from the model. The summary of the fixed effects is presented in Table 4.12.

Table 4. 12: Summary of the fixed effects of time and condition across groups for accentuation

Effect	df	Chi-square	P value	Sig.
time	2	11.28	<0.01	**
ssgroup	2	36.97	<0.0001	***
condition	4	20.59	<0.001	***
time:ssgroup	4	95.78	<0.0001	***

ssgroup had a significant impact on the overall judgement of intonation meanings cued by accentuation. However, this impact was conditioned by *time*, as the interaction term was significant with a very small p value ($\chi^2(4)=95.78$, $p<0.0001$), which implied that the difference between Chinese groups depended on when they took the test, before, after, or two months after the training sessions. *Condition* alone had a main effect on the overall judgement regardless of which group the participants were in and which time point the test was done.

The story was slightly different for phrasing. The optimal fitted GLMM was with *time*, *ssgroup*, *condition*, the interaction of *time* and *ssgroup*, and the interaction of *time* and *condition* as fixed factors. Random effects were of the same structure as for accentuation, i,e. by-subject and by-item adjustment to the intercept, by-item varying slopes for *time*. The two-way interaction of *ssgroup* and *condition* again failed to contribute to the model fit ($\chi^2(8)=7.58$, $p=0.48$), and so did the three-way interaction of *time*, *ssgroup* and *condition* ($\chi^2(16)=13.82$, $p=0.61$). Thus both interaction terms were excluded. The complete main effect of each fixed factor is summarized in Table 4.13.

Table 4. 13: Summary of the fixed effects of time and condition across groups for phrasing

Effect	df	Chi-square	P value	Sig.
time	2	7.80	0.02	*
ssgroup	2	34.93	<0.0001	***
condition	4	74.39	<0.0001	***
time:ssgroup	4	49.51	<0.0001	***
time:condition	8	46.79	<0.0001	***

All effects in the model were significanct Time had a weaker effect for phrasing ($\chi^2(2)=7.80$, $p<0.05$) than it did for accentuation ($\chi^2(2)=11.28$, $p<0.01$), and this effect was conditioned by

ssgroup, as the effect of the interaction between *time* and *ssgroup* was $\chi^2(4)=49.51$, $p<0.0001$, which suggests that the difference between Chinese groups in comprehending phrasing contrasts varied across *time*. A significant effect was also found for *condition* ($\chi^2(4)=74.39$, $p<0.0001$) and this effect was much stronger than it was for accentuation ($\chi^2(4)=20.59$, $p<0.001$). In addition, the interaction between *time* and *condition* also contributed to the explanation of the overall judgement of phrasing, because the effect of this interaction term was significant $\chi^2(4)=46.79$, $p<0.0001$). This indicates that, unlike the effect of *condition,* which remained consistent across *time* and *group* in predicting the comprehensibility of accentuation, the effect of condition was differentiated across *time* for participants' overall judgement of intonation meanings conveyed by phrasing.

The comprehension of tonal contrasts seemed fairly similar to that of accentuation, as the GLMM structure was identical to the one for accentuation but applied to a different dataset. The fixed effects structure was constructed by *time*, *ssgroup*, *condition*, and the interaction of *time* and *ssgroup*. The inclusion of two-way interaction of *time* and *condition* ($\chi^2(4)=3.55$, $p=0.47$)*, ssgroup* and *condition* ($\chi^2(8)=2.81$, $p=0.95$)*,* and three-way interaction of *time*, *ssgroup* and *condition* ($\chi^2(8)=6.00$, $p=0.65$) made the model less fitted to the data. Thus, these three interaction terms were removed from the final model. Random structure was the same as the previous ones, i.e. by-subject and by-item varying intercepts and by-item random slopes for *time*. Table 4.14 presents the statistical results for these interactions.

Table 4. 14: Summary of the fixed effects of time and condition across groups for tone

Effect	df	Chi-square	P value	Sig.
time	2	6.17	0.05	*
ssgroup	2	26.28	<0.0001	***
condition	4	15.00	0.005	**
time:ssgroup	4	33.39	<0.0001	***

Results show that the interaction of *time* and *ssgroup* exerted a significant effect on learners' comprehension of tonal contrasted meanings ($\chi^2(4)=33.39$, $p<0.0001$), as it did for accentuation and phrasing. *Time* and *ssgroup* on its own again played a significant role. This suggests that learners' reaction to tonal contrasts diverged according to which group they belonged to and in which test they participated. The effect of *condition* across time and groups for the judgement of tone was the same ($\chi^2(4)=15$, $p<0.01$), and this effect was less significant than it was for accentuation ($\chi^2(4)=20.59$, $p<0.001$) and phrasing ($\chi^2(4)=74.39$, $p<0.0001$). It suggested that the difference between conditions was the smallest for tonal contrasts, the largest for contrastive phrasing patterns.

Figure 4.5 plots the observed mean accuracy rate for three groups as a function of time. It was already established from the main GLMM that *feature* was not found to be effective in accounting for the judgment of intonation meanings for Chinese participants, and it did not interact with *time* and *ssgroup*. This suggests that there was hardly a difference between learners' performance on accentuation, phrasing and tone. Group differences seemed to occur over time, but these differences tended to be similar across features.

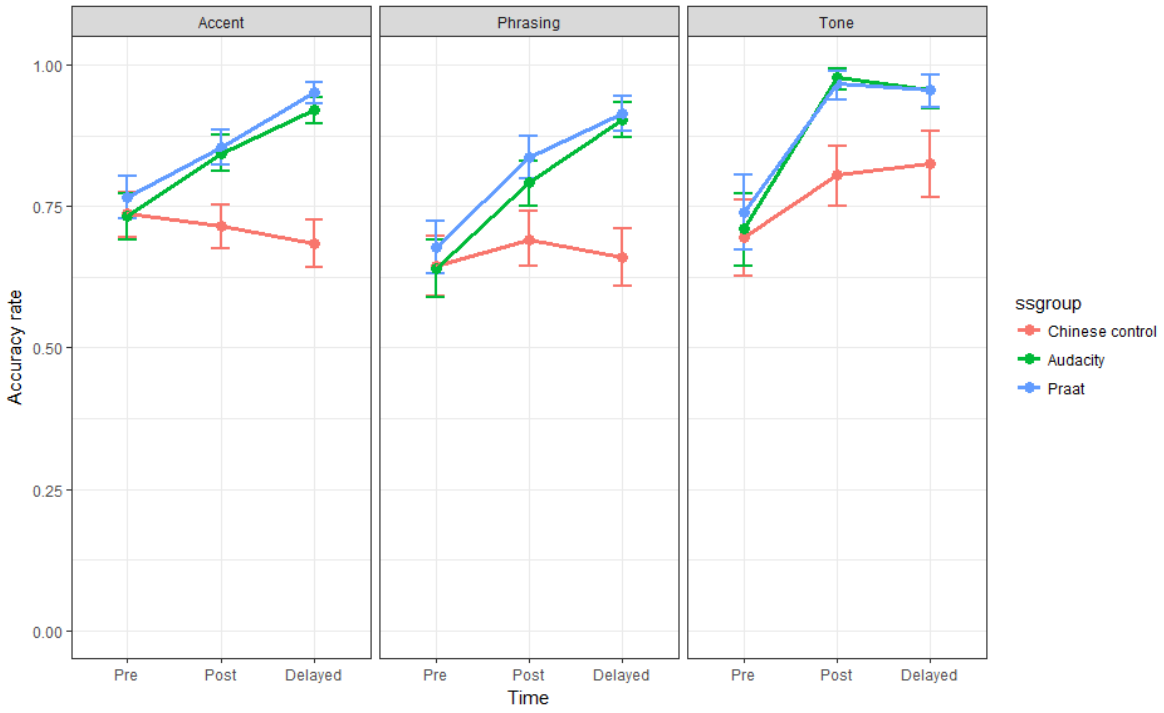


Figure 4. 5: Mean accuracy rate of Chinese groups as a function of time (averaged by condition)

As can be gleaned from Figure 4.5, the three groups roughly started from the same starting point for each feature, as both of the mean accuracy rates and error bars almost overlapped. Over time however, the performance of the treatment groups (the blue and green line) and the control group (the red line) started to deviate and the error bars got larger. Specifically, the performance of the control group on accentuation contrasts decreased in accuracy between the pre- and the delayed post-test, while the performance of the treatment groups improved, reaching more than 90% accuracy. Phrasing seemed to be the hardest among the three features for Chinese participants at the beginning, and it remained difficult for the control group throughout as their mean accuracy rate for each time point remained at around 65%. The treatment groups, on the contrary, increased their sensitivity to contrastive phrasing patterns, reaching a correct rate of about 90% on the delayed post-test. Tonal performance improved slightly for the control group across time, but the scale of the improvement was smaller than that of treatment groups, as the mean accuracy rate and the error bars of the treatment groups did not overlap at all with those of the control group.

Post-hoc comparisons were done by lsmeans () package (Lenth, 2016) in R to discover how big the difference was between groups across time and within each feature. Thus the least-squares means used for comparisons were generated from the sub-GLMM for each feature.

Table 4.15 displays the results of the group differences for performing accentual contrasts within each time point. As mentioned in the previous section, it was apparent that at the beginning of this research program, participants from each group were of similar capability in their comprehension of intonational meanings delivered by accentuations. However, as time passed, this capability developed differently across groups. Significant differences were found between the control group and the treatment groups for post-test (the difference between the control and Audacity was ß=-0.9, $p<0.001$; between the control and Praat was ß=-1.01, $p<0.0001$) as well as for delayed post-test (the difference between the control and Audacity was ß=-1.92, $p<0.0001$; between the control and Praat was ß=-2.50, $p<0.0001$). The difference between the control and treatment groups was larger in the delayed post-test than in the post-test, meaning that learners participating in the training sessions had continued to develop their sensitivity to tonal contrasts as time passed. The two treatment groups developed, however, in a similar fashion, because there was no significant difference between them in the immediate post-test, nor in the delayed post-test.

Table 4. 15: Comparisons of predicted mean accuracy for accentuation within each time (averaged by condition)

Time	Contrast	Estimate	SE	z ratio	p value	Sig.
Pre	Chinese control - Audacity	0.04	0.21	0.20	0.98	N/A
	Chinese control – Praat	-0.16	0.21	-0.78	0.71	N/A
	Audacity – Praat	-0.19	0.21	-0.98	0.59	N/A
Post	Chinese control - Audacity	-0.90	0.23	-3.98	<0.001	***
	Chinese control – Praat	-1.01	0.23	-4.43	<0.0001	***
	Audacity – Praat	-0.11	0.23	-0.47	0.88	N/A

Delayed post	Chinese control - Audacity	-1.92	0.25	-7.55	<0.0001	***
	Chinese control – Praat	-2.50	0.29	-8.43	<0.0001	***
	Audacity – Praat	-0.58	0.32	-1.79	0.17	N/A

(NB: results were given on the log odds ratio scale, not the response scale. P value was adjusted by Tukey method for comparing a family of 3 estimates)

The development of each group across time is presented in table 4.16. It can be seen that the probability of getting a correct answer was not improved at all across time for the control group, as none of the comparisons was found to be significant. For the Audacity group, the chance of making a correct judgement in pre-test was marginally lower than in post-test (ß=-0.89, p=0.07), but was significantly lower than in delayed post-test (ß=-1.7, p<0.001). The correct rate of post-test was also significantly worse than that of delayed post-test (ß=-0.81, p<0.01). This revealed that the Audacity group improved continuously across time, as the correct rate for the task was always higher than at the previous time point. The Praat group also improved through time, but the improvement from pre- to post-test was not significant, ß=-0.8, p=0.12, whereas the delayed test showed a dramatic improvement from pre-test (ß=-2.07, p<0.0001), and so did the improvement from post-test (ß=-1.28, p<0.0001). Briefly, both treatment groups continued to develop their skills in comprehending tonal contrasts post-training.

Table 4. 16: Comparisons of predicted mean accuracy for accentuation within group across time (averaged by condition)

Group	Contrast	Estimate	SE	z ratio	p value	Sig.
Chinese control	Pre - Post	0.05	0.39	0.13	0.99	N/A
	Pre - Delayed	0.26	0.40	0.65	0.79	N/A
	Post - Delayed	0.21	0.18	1.18	0.46	N/A

Audacity	Pre - Post	-0.89	0.40	-2.23	0.07	.
	Pre - Delayed	-1.70	0.42	-3.98	<0.001	***
	Post - Delayed	-0.81	0.24	-3.32	<0.01	**
Praat	Pre - Post	-0.80	0.40	-1.98	0.12	N/A
	Pre - Delayed	-2.07	0.46	-4.55	<0.0001	***
	Post - Delayed	-1.28	0.29	-4.40	<0.0001	***

(NB: results were given on the log odds ratio scale, not the response scale. P value was adjusted by Tukey method for comparing a family of 3 estimates)

Figure 4.6 presents the predicted mean accuracy rates of doing contrastive accentuation for all groups across time. The difference between the control group and the treatment groups, and the similarity between two treatment groups, along with the way in which each group developed through time were now much more trackable. The control group remained at about 70% correct rate at each time point, while the treatment groups increased their correct rate from around 75% to 90% and then to more than 95%. The slope of the change between pre- and post-test for the Audacity group (the pink line) was steeper than that for the Praat group (the green line), resulting in differences for the improvement for two groups.

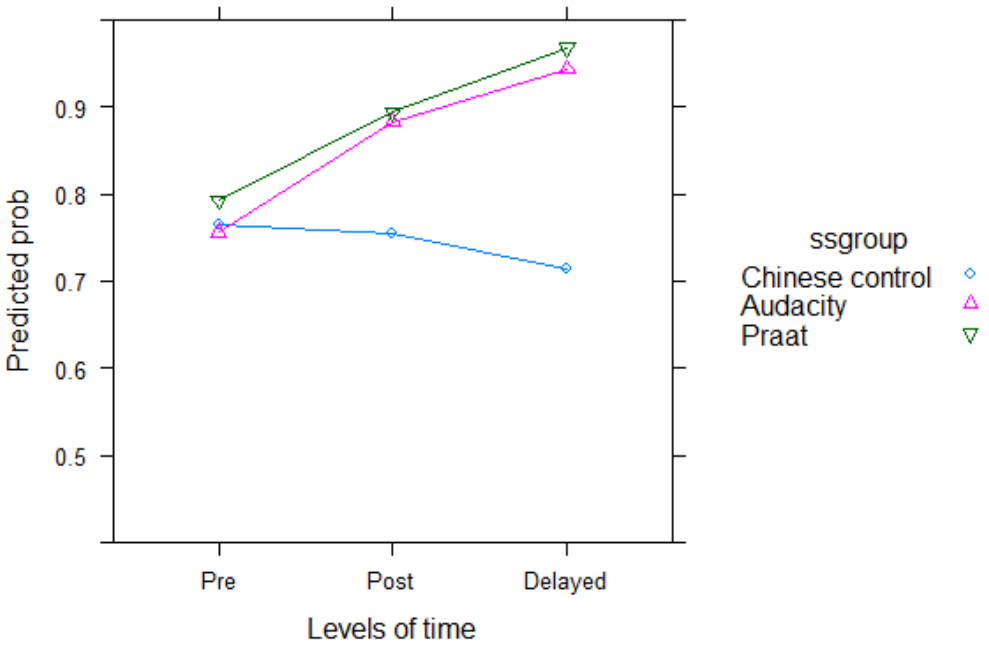


Figure 4. 6: Predicted mean accuracy of accentuation for Chinese groups across time

The same pairs of comparisons were made for phrasing, as shown in table 4.17 and 4.18. Also with the display of the predicted mean accuracy rate for all groups across time in figure 4.7, it can be concluded that the control group again rarely changed in their comprehension of phrasing contrasts from pre- to delayed post-test, with a consistent 75% correct rate; the difference between their pre- and post-test was almost non-identifiable, ß=-0.09, p=0.99, and so was the difference between their pre- and delayed post-test, ß=0.12, p=0.97. Predictably, there was no significant difference between their post- and delayed post-test, ß=0.21, p=0.67.

The two treatment groups, with a similar starting level of accuracy rate to the control group (confirmed by the non-significant difference of the pre-test between the three groups), increased their sensitivity to the meanings cued by phrasing over time; the difference between the control and treatment groups therefore appeared after the pre-test, which was evidenced by the significant results of the comparisons for post-test between the control and Audacity group (ß=-0.6, p=0.01) and between the control and Praat group (ß=-0.93, p=0.0001). The difference of the delayed post-test between the control and treatment groups were even larger

than for the post-test, as the predicted accuracy rate for the delayed post-test for both Audacity and Praat group reached up to 95%±2% SE. No significant difference was uncovered between Audacity and Praat group for the test at any time point.

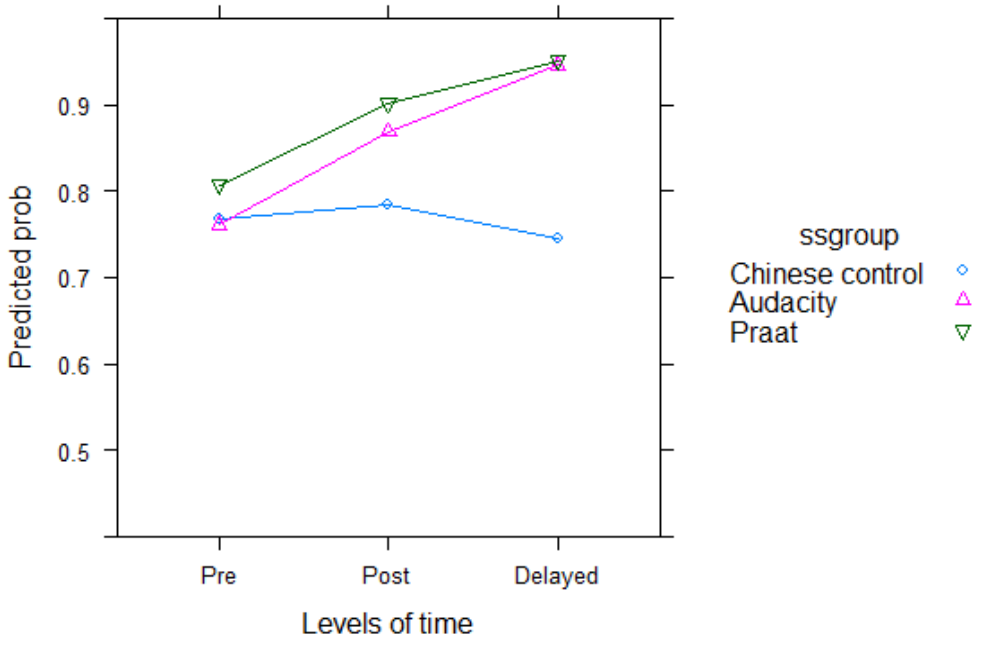


Figure 4. 7: predicted mean accuracy of phrasing for Chinese groups across time

Table 4. 17: Comparisons of predicted mean accuracy for phrasing within each time (averaged by condition)

Time	Contrast	Estimate	SE	z ratio	p value	Sig.
Pre	Chinese control - Audacity	0.04	0.22	0.17	0.98	N/A
	Chinese control - Praat	-0.23	0.22	-1.10	0.54	N/A
	Audacity - Praat	-0.27	0.22	-1.23	0.44	N/A
Post	Chinese control - Audacity	-0.60	0.21	-2.88	0.01	**
	Chinese control - Praat	-0.93	0.22	-4.26	0.0001	***
	Audacity - Praat	-0.32	0.23	-1.44	0.32	N/A

Delayed post	Chinese control - Audacity	-1.79	0.25	-7.19	<0.0001	***
	Chinese control - Praat	-1.90	0.26	-7.19	<0.0001	***
	Audacity - Praat	-0.11	0.29	-0.38	0.92	N/A

(NB: results were given on the log odds ratio scale, not the response scale. P value was adjusted by Tukey method for comparing a family of 3 estimates)

It should be noted that the improvement from pre- to post-test was non-significant for both Audacity (ß=-0.74, p=0.44) and Praat group (ß=-0.79, p=0.4), even though visual inspection of figure 4.7 shows an upward trend. Both groups did, however, exhibit a significant improvement when performing the delayed post-test, as the difference between pre- and delayed post-test for Audacity was ß=-1.7, p=0.005, and for Praat it was ß=-1.55, p=0.01, and the difference between post- and delayed post-test for Audacity was ß=-0.97, p=0.004, and for Praat it was ß=-0.76, p=0.05. In both cases (from pre- to post-test, and from post- to delayed post-test), the Audacity group tended to show bigger improvement than the Praat group.

Table 4. 18: Comparisons of predicted mean accuracy for phrasing within group across time (averaged by condition)

Group	Contrast	Estimate	SE	z ratio	p value	Sig.
Chinese control	Pre - Post	-0.09	0.60	-0.16	0.99	N/A
	Pre - Delayed	0.12	0.52	0.23	0.97	N/A
	Post - Delayed	0.21	0.25	0.86	0.67	N/A
Audacity	Pre - Post	-0.74	0.60	-1.22	0.44	N/A

	Pre - Delayed	-1.70	0.54	-3.15	0.005	**
	Post - Delayed	-0.97	0.30	-3.20	0.004	**
Praat	Pre - Post	-0.79	0.61	-1.30	0.40	N/A
	Pre - Delayed	-1.55	0.55	-2.81	0.01	**
	Post - Delayed	-0.76	0.32	-2.35	0.05	*

(NB: results were given on the log odds ratio scale, not the response scale. P value was adjusted by Tukey method for comparing a family of 3 estimates)

Figure 4.8 shows the predicted mean accuracy rates across time with regard to tonal contrasts. Unlike accentuation and phrasing, tonal contrasts seemed to improve for all three groups from pre- to post-test. However, statistical analysis showed that this improvement was not significant for the control group, ß=-0.91, p=0.31, while it was for both the Audacity (ß=-3.44, p<0.0001) and the Praat group (ß=-2.84, p<0.001). The difference between post- and delayed post-test for the control group was as minor as that for the Audacity and Praat group, and this was verified by pairwise comparisons of the tests within each group: the difference was not significant for all three groups (for the control group, ß=-0.02, p=1.00, for Audacity group, ß=0.9, p=0.36, for Praat group, ß=0.49, p=0.71). This demonstrated that two treatment groups developed their comprehension of tonal contrasts to a nativelike level with the three-week intensive training sessions (99%±1% SE correct rate for Audacity group and 98%±1% SE for Praat group) and they retained their improved ability for at least two months (similar correct rates were obtained on the delayed post-test). The full details of the comparison results of the time course of the change within groups are presented in table 4.20.

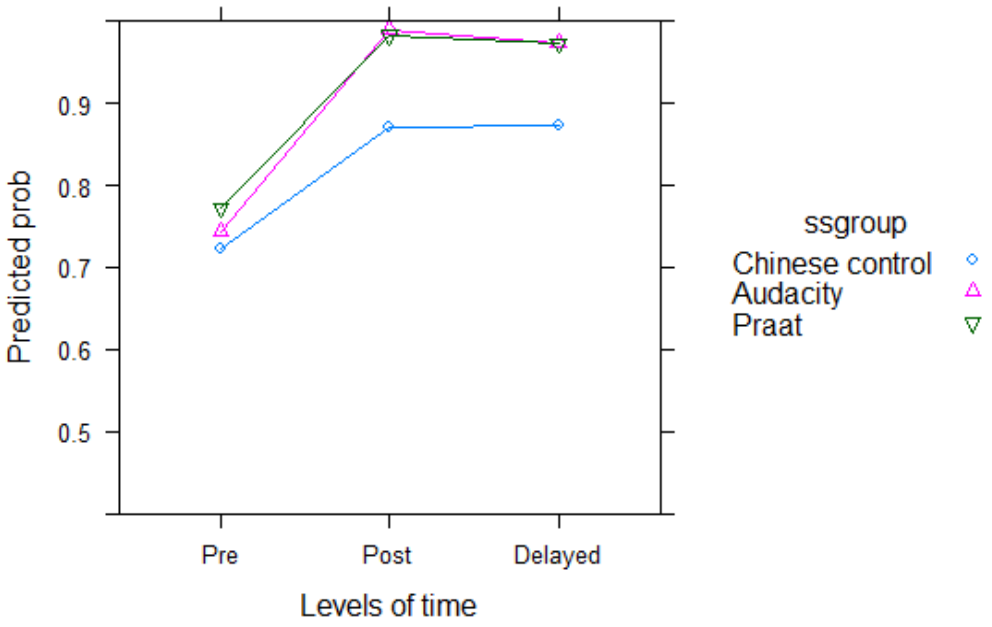


Figure 4. 8: Predicted mean accuracy of tone for Chinese groups across time

Table 4. 19: Comparisons of predicted mean accuracy for tone within each time (averaged by condition)

Time	Contrast	Estimate	SE	z ratio	p value	Sig.
Pre	Chinese control - Audacity	-0.10	0.32	-0.33	0.94	N/A
	Chinese control - Praat	-0.26	0.32	-0.81	0.70	N/A
	Audacity - Praat	-0.16	0.32	-0.48	0.88	N/A
Post	Chinese control - Audacity	-2.60	0.58	-4.50	<0.0001	***
	Chinese control - Praat	-2.16	0.51	-4.28	0.0001	***
	Audacity - Praat	0.44	0.68	0.65	0.79	N/A
Delayed post	Chinese control - Audacity	-1.69	0.47	-3.61	0.0009	***
	Chinese control - Praat	-1.66	0.49	-3.41	0.0019	**
	Audacity - Praat	0.03	0.57	0.05	1.00	N/A

(NB: results were given on the log odds ratio scale, not the response scale. P value was adjusted by Tukey method for comparing a family of 3 estimates)

Cross-sectional comparisons of group differences, as shown by table 4.19, reveal that the control group was of similar capability to the treatment groups in the comprehension of intonational meanings contrasted by nuclear tones at the beginning, as no significant difference was found between them at pre-test. The intonation training, as hypothesized, improved this ability for participants who were trained over a three-week period, and the control group was left behind the Audacity and Praat group in terms of their performance in the post- and delayed post-test (in the post-test, the difference between the control and Audacity was ß=-2.6, $p<0.0001$; between the control and Praat was ß=-2.16, $p=0.0001$; in the delayed post-test, the difference between the control and Audacity was ß=-1.69, $p<0.001$; between the control and Praat was ß=-1.66, $p<0.01$). Different self-directed practice platforms (auditory-only/audio-visual), however, did not lead to a difference in performance, with a non-significant results for the comparisons between Audacity and Praat group in both post- (ß=0.44, $p=0.79$) and delayed post-test (ß=0.03, $p=1$).

Table 4. 20: Comparisons of predicted mean accuracy for tone within group across time (averaged by condition)

Group	Contrast	Estimate	SE	z ratio	p value	Sig.
Chinese control	Pre - Post	-0.94	0.64	-1.47	0.31	N/A
	Pre - Delayed	-0.96	0.50	-1.92	0.13	N/A
	Post - Delayed	-0.02	0.36	-0.06	1.00	N/A
Audacity	Pre - Post	-3.44	0.80	-4.32	<0.0001	***
	Pre - Delayed	-2.54	0.59	-4.33	<0.0001	***
	Post - Delayed	0.90	0.66	1.36	0.360	N/A
Praat	Pre - Post	-2.84	0.74	-3.81	0.0004	***
	Pre - Delayed	-2.36	0.60	-3.91	0.0003	***

	Post - Delayed	0.49	0.62	0.79	0.71	N/A

One more important set of results that cannot be seen from the tables of all pairwise comparisons above emerges from table 4.21, which summaries the predicted mean accuracy rates of all intonational features for the three groups at each testing point. Not only did the treatment groups improve their overall ability to comprehend intonation meanings over time, but they also varied less in performing these tasks across time, as the standard errors of the treatment groups were getting smaller from pre- to delayed post-test, while those of the control group did not decrease consistently. Additionally, the 95% confidence intervals for each training group reveal that the predicted means for the treatment groups were showing more accuracy and more representative than that for the control group; this suggests that Chinese learners of English are essentially able to identify the correct intonation meaning cued by any feature if they are explicitly instructed about the form and meanings of English intonation and they practice for a certain amount of time with delicately designed practice materials, no matter through what medium. Visualised intonation patterns (Praat) did not facilitate performance more than ear training alone (Audacity) in the comprehension tasks. Relatively unchanging wide 95% confidence intervals from pre- to delayed post-test for the control group indicated that the predicted mean accuracy rates for this group tended to be less representative; Chinese learners of English who did not receive any training on intonation performed randomly on comprehension tasks.

Table 4. 21: Predicted mean accuracy rates of all features for three groups across time (averaged by condition)

Feature	Group	Time	Predicted probability	SE	asymp.LCL	asymp.UCL
Accent	**Control**	Pre	0.76	0.04	0.68	0.83
		Post	0.75	0.07	0.61	0.86

		Delayed	0.71	0.07	0.56	0.83
	Audacity	Pre	0.76	0.04	0.67	0.82
		Post	0.88	0.04	0.79	0.94
		Delayed	0.94	0.02	0.89	0.97
	Praat	Pre	0.79	0.04	0.71	0.85
		Post	0.89	0.03	0.81	0.95
		Delayed	0.97	0.01	0.93	0.99
Phrasing	**Control**	Pre	0.77	0.14	0.42	0.94
		Post	0.78	0.06	0.64	0.88
		Delayed	0.75	0.08	0.56	0.87
	Audacity	Pre	0.76	0.14	0.41	0.94
		Post	0.87	0.04	0.76	0.93
		Delayed	0.95	0.02	0.88	0.98
	Praat	Pre	0.81	0.12	0.47	0.95
		Post	0.90	0.03	0.81	0.95
		Delayed	0.95	0.02	0.89	0.98
Tone	**Control**	Pre	0.72	0.09	0.53	0.86
		Post	0.87	0.07	0.66	0.96
		Delayed	0.87	0.06	0.71	0.95
	Audacity	Pre	0.74	0.08	0.56	0.87
		Post	0.99	0.01	0.95	0.99
		Delayed	0.97	0.02	0.92	0.99
	Praat	Pre	0.77	0.08	0.59	0.89
		Post	0.98	0.01	0.93	0.99
		Delayed	0.97	0.02	0.91	0.99

Chapter V : General Discussion

This chapter discusses the results presented in Chapter IV and comprises four sections. The first three sections address the three research questions of this thesis. They are: (1) Can L1-Mandarin Chinese learners of English distinguish the intonation meanings conveyed by contrastive tonality, tonicity, and tone patterns? (2) With the explicit instruction and the self-paced practice, is Chinese learners' perception of intonation meanings improved? (3) Is there any difference of the effects between the auditory and audio-visual feedback on improving Chinese learners' perception of intonation meanings? The fourth section summaries the feedback and comments on the training methods provided to the two Chinese experimental groups.

5.1 L1-Mandarin Chinese EFL learners' perception of English intonation

By analysing the results of the pre-test, it was not surprisingly found that L1-Mandarin Chinese learners in general did significantly worse than the native speakers in all three intonational elements (tonality, tonicity, and tone) that encode distinctive meanings. Although Chinese learners' observed correct rate of any intonation element was above the chance level, it does not necessarily mean that most of the Chinese learners were able to consistently understand intonational meanings, for the fact that a wide range of confidence intervals was found for those predicted mean accuracy rates based on the reliable inferential analyses. It rather indicates that Chinese EFL learners tend to perceive semantic and pragmatic meanings of intonation in a random way. This is in accordance with the findings from Atoye (2005) who examined 120 Nigerian learners' interpretation of intonation meanings. He found that his participants could discriminate the phonetic differences between the paired intonation patterns but could not assign appropriate meanings to these patterns.

In addition, the present study has shown that learners' incompetence in meaning judgement was similarly reflected on 3Ts. Although the mean accuracy rate of tonality was the lowest,

its difference between tonicity and tone was neither significant, which indicates that adult Chinese learners' comprehension ability of intonation features are developed in parallel. This is inconsistent with the findings from previous studies of L1 child prosodic development. Cruttenden (1985) and Crystal (1986) argued that tonicity and tone are developed earlier than tonality for L1-English children both in comprehension and production, but their analyses were based on descriptive statistics, leaving their argument less convincing.

When examining perception ability from learners with different levels of English proficiency, only those with very advanced level (C2 in CEFR) stood out as having a better understanding of meanings cued by tonality with a predicted accuracy rate of 93% ± 4%SE, which was very close to that of native speakers (97% ± 2%SE). Learners with other levels (B2, B1, and C1) by comparison, were similar in perceiving contrastive tonality. However, learners with C2 level of English did not show significant advantages in discriminating meanings of contrastive tonicity and tone, indicating that learners with nativelike English proficiency would still face difficulty in this regard. In other words, English proficiency seemed not to be a decisive factor for individual variation that was found among Chinese learners in doing comprehension tasks as a whole. Learners' ability to decipher intonational meanings tends to be unstable, and there is no guarantee for even the most advanced learners to process the correct meaning for tonicity and tone.

Previous studies which claimed that Chinese learners were capable of perceiving and producing tonicity in a nativelike fashion (Ji *et al.*, 2009; Ding *et al.*, 2012b) were focused on the phonetic dimension of intonation whereby subjects were required to locate the perceived prominence, and they attributed their findings to L1 transfer as both Mandarin Chinese and English employ similar phonetic manipulations to prominence (or accentuation) such as lengthened duration, expanded pitch range, and raised intensity (see section 2.2 for the detailed comparison of Mandarin Chinese and English intonation). The current study, however, did not find that intonational meanings conveyed by contrastive tonicity was easier for Chinese learners to perceive, even for very advanced learners. This finding is closely

aligned with the results from Ortega-Llebaria and Colantoni (2014) who examined learners' perceptual and production ability in both context-free and context-rich tasks. They have found that context-rich tasks such as identification and interpretation of intonational meanings imposed more challenges for L2 learners than context-free tasks such as perception of accentuations, as learners are subject to a higher cognitive load when making judgements of the context-rich tasks than of context-free tasks (Derwing *et al.*, 2012). It is therefore predictable that L1-Mandarin Chinese learners would encounter difficulty in distinguishing contrastive intonational meanings via different accentuation locations, although Mandarin Chinese resembles English in terms of phonetic realisations of accentuation and Chinese learners might be sensitive to these phonetic cues.

In Mennen's (2015) working theory of L2 intonation acquisition (LILt), learners' intonation ability is accounted not just by phonetic and phonological dimensions, but also by semantic dimension, and it predicts that the development of L2 intonation skills might vary in these dimensions. The present study, alongside Ortega-Llebaria and Colantoni (2014) have provided evidence that adult learners of English do have difficulty in doing context-based intonation tasks, though learners' L1 and the target L2/FL are thought to be similar along the phonetic dimension of accentuation and thus it should not be a problem for doing tasks phonetically-driven as in Ji *et al.* (2009) and Ding *et al.* (2012b).

It has been evidenced by previous literature that learners with higher overall English proficiency would exhibit nativelike manipulation of phonetic cues of 3Ts (Wang *et al.*, 2011a) and nativelike identification of accentuation placement to native speakers (Wang *et al.*, 2010a), supporting the argument that the acquisition of L2 intonation system is a gradual process (Trofimovich and Baker, 2006; He *et al*., 2012), just like the acquisition of L2 segments predicted by SLM (Flege, 1995) and PAM-L2 (Best and Tyler, 2007). This study, however, has not found any positive correlation between Chinese adult EFL learners' overall English proficiency and their perceptual ability of intonation meanings, and if L2 production is perception-oriented as predicted by SLM and PAM-L2 for segments and LILt for

intonation, Chinese learners' production of semantically meaningful English utterances would be a reflection of their comprehension ability, which is not correlated to their overall English proficiency. Therefore, I argue that L2 or foreign language learners' perception of intonation meaning is not reflected on their gradually improved English proficiency as evidenced in this thesis. The existing working model on L2 intonation acquisition (i.e. LILt) should include perception as well, and the predictions of the relationship between perception and production of L2 intonation should be based on more extensive research evidence.

5.2 The effects of the explicit instruction and self-paced intonation practice

As for training effects, it is as anticipated that Chinese learners who receive explicit instruction and accompanying practice will develop their understanding of intonational meanings encrypted by tonality, tonicity, and tone. More importantly, this effect seems to persist as the difference between the delayed post-test and the immediate post-test was significant for the treatment groups in terms of tonality and tonicity, whereas a non-significant result for tone was found because of the ceiling effect (the results on the immediate post-test almost reached the native speakers' level). Training studies that failed to yield a positive effect for instruction such as in Kurt *et al.* (2014) may be partially because of lacking enough input and practice. The instructed patterns, though as they were claimed to be meaningful, were restricted to three sentence types—wh-question, tag question, and statements, which were heavily questioned on their pedagogical worthiness. Yang (2010) even found a negative effect of "listen-and- repeat" training on Chinese learners' production of English intonation contours; the results from the first test was even worse than the pre-test. Both studies have ignored the importance of communicative core of intonation. The teaching of semantics and pragmatics of intonation is as crucial as teaching its phonetic and phonological system (He *et al*., 2010; 2012), as Lord (2010) points out that explicit instruction can make learners aware of the use of particular pronunciation features in the target L2, and as a result accelerate their acquisition of those features. The present study has proven that after being equipped with

intonation knowledge, Chinese adult learners were more sensitive to intonation patterns and able to consistently make correct choice of intonation meanings than their peers who did not receive any training.

The findings of the long-term effect of the intonation training is very promising and further rationalise the effectiveness of the instruction and practice materials which were theoretically and pedagogically referred to. The reliability of instruction materials is often questioned by applied linguists on being lack of theoretical stance for instruction itself (Thomson and Derwing, 2014), as well as for the failure of applying theoretically-oriented experimental findings into the materials (Rogerson-Revell, 2011). The present study has made a successful attempt to combine theories and empirical evidence into training materials that are specifically designed for adult L1-Chinese learners of English. As reviewed in section 1.1, the pronunciation textbooks for college students on the Chinese market are not just being criticized for depriving of pedagogical justifications, but also for adopting outdated intonation theories, such as over-simplified instruction of intonation contours in Kurt *et al.* (2014) as mentioned above. In short, this study supports the view that explicit instruction of intonation form and functions is beneficial for adult learners' perception of intonation meanings as long as the instruction materials are theoretically and pedagogically founded. Having the relevant knowledge of intonation form and functions will enhance self-learning and in turn bring further development of the comprehension ability of intonation of the target L2 or foreign language.

5.3 Auditory vs. audio-visual intonation training

By comparing the results of the post- and delayed post-test between the *Praat* and *Auditory* group, no significant difference was found in both tests, and in fact the *Praat* group progressed in an identical path as the *Audacity* group, indicating that visualised pitch tracings were not found to be of additional help for learners' comprehension of intonation meanings.

Previous studies (Gorgian et al., 2013; Wilson, 2008) which found a positive effect of using Praat focused on *production*, and as many other studies using computer-based visualisation tools reviewed in section 2.4 (e.g. Anderson-Hsieh, 1992; Hardison, 2005), they failed to include a control group who could only access auditory feedback in their experiment. The visualised intonation is audio-visual per se, so it is uncertain whether the observed improvement reported in these studies can be entirely attributed to the visualisation of intonation. Empirical research of auditory training has found that L2-English learners with different L1 backgrounds can benefit from the CPR (Cued Pronunciation Recording) practice method in perceiving pause, word stress, and sentence-final tone and in producing word and sentence stress (Tanner and Landon, 2009).

In Tanner and Landon (2009), the learning targets were briefly but explicitly instructed for 60 minutes to 75 participants. Then they were assigned to two groups, one with 10-minute self-directed CPR training daily for the following 11 weeks, the other left with no homework. The experimental group practiced the learning targets by listening to the recordings produced by native speakers and noting down the targets on the written materials. Then they produced and recorded their own production based on their notes for later comparisons to the native samples. The Audacity group in the present study went through a similar practice process as the participants in Tanner and Landon's (2009) study, and it turns out that Chinese EFL learners' comprehension of intonation meanings was indeed greatly improved via auditory training.

It is not surprising the *Praat* group did not show greater improvement in perception of intonation meanings, for the fact that contrastive patterns signalling different meanings were full of phonetic variations which were auditorily detectable. The practice materials and the process of the *Praat* group was primarily the same as the *Audacity* group, except that all the recordings were visually accessible. Though useful acoustic features associated with intonation was taught before the first session of practice, it was observed that sometimes the redundant visualised pitch curves caused by the octave errors, the background noise, and even

other participants' sounds might frustrate the individual who was comparing his/her own production to the native patterns, as in such cases, his/her own pitch curves looked different from the provided samples and sometimes even looked nothing like the samples. Although the researcher encouraged them to raise their hand and ask for help, some of them preferred observing the native samples and producing their own without comparison between the two. de Bot (1983) found that the auditory group in his study spent more time on listening to native recordings and less time on oral practicing than the audio-visual group. In the present study, on the contrary, the *Audacity* group tended to spend more time on oral practicing and comparison of their own production and the provided samples, while the *Praat* group spent most of their time on observing and listening to the native samples and less time on oral practicing.

All in all, the present study has provided robust evidence for the effectiveness of the auditory training on adult Chinese EFL learners' perception of intonation meanings. Additional usefulness and effectiveness of audio-visual training, however, has not been justified, indicating that the auditory training can be as beneficial as the audio-visual training at least in terms of the comprehension of intonation.

5.4 Chinese learners' feedback on the intonation training method

Based on the results from Part 1 of the post-test questionnaire, more than 90% of participants evaluated themselves as knowing how to manipulate tonality, tonicity, and tone, compared to less than 20% in the pre-test questionnaire. When asking if they can properly perceive the 3Ts, more than 60% said yes and more than 30% said they were very sure that they can. 32 out of 40 participants (80%) expressed their confidence in using proper intonation, which was a huge increase than in the pre-test questionnaire (less than 10%). More than 90% said that they now consciously pay attention to intonation whenever they hear or speak English. To sum up, the intonation training provided to these participants have raised their awareness of

the knowledge of English intonation and developed their confidence with their intonation ability.

Part 2 of the questionnaire elicited participants' comments and feedback on the training methods. All participants regardless of the group they were in (*Praat* or *Audacity*) approved that the training course provided by the researcher was useful for their future self-improvement in English intonation (73% chose quite useful, and 27% chose extremely useful). For the *Audacity* group, about 95% of the participants thought that the method they were trained with was useful, compared to 80% of the participants for the *Praat* group. When asked how happy they were with the English models used in the training course, 76% were very happy, and 37% hoped for more real-life materials, and about 18% looked for more examples of other English varieties. When they were asked if they were about to teach English intonation to their future students, more than 88% answered yes. But some participants also expressed their concerns of teaching intonation because of the complexity of the whole system and the time it will take to take in all the information. In general, the feedback on the training course was positive as reviewed by the involved participants, and some participants told the researcher that they wished they could have attended the training course earlier or more training courses could be provided.

Chapter VI: Conclusion

This chapter summarises the major findings of this thesis and provides pedagogical implications for intonation teaching to adult Chinese EFL learners, followed by an acknowledgement of the limitations of the present study and some suggestions for future research.

This thesis was conducted to address three major research questions. The first question was satisfactorily addressed by analysing the comprehension task from the pre-test. Chinese learners were significantly worse than native speakers in identifying semantic and pragmatic meanings of English intonation encoded into accentuation, phrasing and tone. They were also found to have less agreement over their choices as the statistical analysis revealed that their judgements of intonation meanings were more randomly distributed. Their difficulties in understanding the three essential intonation elements were equally scaled. Chinese learners with the most advanced level of English (nearly nativelike) exhibited greater capability of understanding tonality-contrastive patterns, while their ability to decipher tonicity and tone was similar to less advanced learners. This thesis made a first attempt to examine Chinese EFL learners' comprehension of English intonation and showed that even for proficient students, understanding the intonation meaning was a challenging task. Although there is a great deal of research which has found that Chinese students are able to perceive and produce native-like accentuation, their understanding of accentuation-induced meanings was not as good as anticipated, indicating that a further and deeper investigation is urgently needed for knowing the underlying causes of learners' advantages and disadvantages at the phonetic and phonological levels of intonation. Understanding their comprehension ability might be a right direction.

The second and third questions were addressed by the discussion of the results from the post- and delayed post-test. The tailor-made instruction and practice materials were confirmed to be effective, as both treatment groups were seen to have a significant improvement on their performance of the post-test, while the further improvement on the delayed post-test suggests

that by means of explicit instruction with specifically designed practice activities, learners' knowledge of intonation was retained and enhanced on their own after they finished the training. Audio-visual feedback on learners' acquisition of intonation might be not as facilitating as it was claimed, at least not for the understanding of intonation meanings. All in all, intonation training seems to be useful and plays a strong intervening role to help raise learners' awareness of intonation. Once they start to pay attention to intonation, the training effect might expand over a longer time.

For the pedagogical implications of intonation teaching, this study strongly supports that intonation training should concentrate on the semantic and pragmatic knowledge rather than pure imitations of patterns (He *et al.*, 2010; Saalfeld, 2011; Mitrofanova, 2012). In addition, intonational features, particularly accentuation and tone deserve to be taught as learners are less likely to naturally acquire them. Phrasing, on the other hand, though preferably comprehended by the most proficient learners, still needs intervention to the researcher's knowledge, because most of the Chinese learners have a very limited access to considerable amount of native input (Zhang, 2015) which plays a crucial role for L2 phonology acquisition (Johnson, 1997). Without intervention, they would not even notice these features in which even with a subtle change in its form, the meaning they deliver would be thoroughly different. Considering that intonation itself is complex and indeed very difficult for L2 learners to acquire, intonation training should start with the simplest context-free patterns, such as nuclear tone categories (falling, rising, and falling-rising), and gradually move to richer context-based features, such as tonicity (placement of the accentuation) and tonality.

The limitations of this study are two-fold. Firstly, the number of male participants was very small, and participants' daily exposure to English was not controlled for. Further research can be done by controlling for gender and exposure of the target L2 to explore learner variation in perception of intonation meanings. Secondly, the length of each intonation training session

was a bit long. Intonation training in practical situations can be shorter each time (about 30 minutes) but last longer, for example, over the whole academic year.

Other suggestions for the future research are: (1) production can be targeted to see if there is any difference of the effects between auditory and audio-visual training on production of meaningful intonation, (2) exploring the relationship between the semantic and phonetic/phonological dimensions of L2/FL intonation to further develop the L2 intonation acquisition theory (e.g. LILt).

References

't Hart, J., Collier, R. and Cohen, A. (1990) *A perceptual study of intonation: an experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.

Altmann, H. (2006) *The perception and production of second language stress; A cross-linguistic experimental study*. Doctoral thesis. University of Delaware, Newark.

Anderson-Hsieh, J. (1992) 'Using electronic visual feedback to teach suprasegmentals', *System*, 20(1), pp. 51-62.

Anderson-Hsieh, J. (1994) 'Interpreting visual feedback on English suprasegmentals in computer assisted pronunciation instruction', *CALICO Journal*, 11(4), pp. 5-22.

Anderson-Hsieh, J. and Venkatagiri, H. (1994) 'Syllable duration and pausing in the speech of Chinese ESL speakers', *TESOL Quarterly*, 28(4), pp. 807-812.

Atoye, R. (2005) 'Non-Native perception and interpretation of English intonation', *Nordic Journal of African Studies*, 14(1), pp. 26-42.

Audacity Team (2014). Audacity(R): Free Audio Editor and Recorder [Computer program]. Version 2.0.0

Baayen, H., Bates, D., Kliegl, R. and Vasishth, S. (2015) *RePsychLing: Data sets from Psychology and Linguistics experiments. R package version 0.0.4.*

Baayen, R.H. (2008) *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Barr, D.J., Levy, R., Scheepers, C. and Tily, H.J. (2013) 'Random effects structure for confirmatory hypothesis testing: Keep it maximal', *Journal of Memory and Language*, 68(3), pp. 255-278.

Bates, D., Kliegl, R., Vasishth, S. and Baayen, H. (2015a) 'Parsimonious mixed models', *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015b) 'Fitting Linear Mixed-Effects Models Using lme4', *Journal of Statistical Software*, 67(1), pp. 1-48.

Beckman, M. and Ayers, G. (1997) 'Guidelines for ToBI labelling', *The OSU Research Foundation*, 3.

Best, C.T. (1995) 'A Direct Realist View of Cross-Language Speech Perception', in Strange, W. (ed.) *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Baltimore: York Press, pp. 171-204.

Best, C.T. and Tyler, M.D. (2007) 'Nonnative and second-language speech perception: Commonalities and complementarities', in Munro, M.J. and Bohn, O.S. (eds.) *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, pp. 13-34.

Bi, Y., Ganushchak, L.Y., Konopka, A.E., Ren, G., Sui, X. and Chen, Y. (2016) 'Prosodic encoding of information structure in Mandarin Chinese: Evidence from picture description task', *Speech Prosody*. Boston.

Boersma, Paul & Weenink, David (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1, retrieved 13 July 2019 from http://www.praat.org/

Bolinger, D. (1968) *Aspects of Language*. New York: Harcourt Brace Jovanovich.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. and White, J.-S.S. (2009) 'Generalized linear mixed models: a practical guide for ecology and evolution', *Trends in Ecology & Evolution*, 24(3), pp. 127-135.

Brazil, D. (1997) *The Communicative Value of Intonation in English*. Cambridge: Cambridge University Press.

Brazil, D., Coulthard, M. and Johns, C. (1980) *Discourse Intonation and Language Teaching*. Harlow: Longman.

Brown, G., Currie, K.L. and Kenworthy, J. (1980) *Questions of Intonation*. London: Helm.

Bu, Y. (2003) *English Intonation: Its Form, Function and Application*. Beijing: Foreign Language Teaching&Research Press.

Busà, M.G. (2008) 'New perspectives in teaching pronunciation', in Baldry, A., Pavesi, M., Torsello, C.T. and C.Taylor (eds.) *From Didactas to Ecolingua: an Ongoing Research Project on Translation and Corpus Linguistics*. EUT - Edizioni Università di Trieste. Available at: http://hdl.handle.net/10077/2850.

Cao, J. (2000) 'Intonation structure of spoken Chinese: Universality and characteristics', *From Traditional Phonology to Modern Speech Processing.–Beijing: Institute of Linguistics CASS*.

Cao, J. (2004) 'Intonation structure of spoken Chinese: Universality and specificity', in Fant, G., Fujisaki, H., Cao, J. and Xu, Y. (eds.) *From Traditional Phonology to Modern Speech Processing: Festschrift for Professor Wu Zongji's 95th Birthday*. Beijing: Foreign Language Teaching and Research Press, pp. 31-38.

Cao, J. (2005) 'Yinduan yangchang de butong leixing jiqi yun lv jiazhi [Types of segmental lengthening and its prosodic indication]', *Nanjing shifan Daxue Wenxueyuan Xuebao [Report of the School of Literature of Nanjing Normal University]*, 4.

Cao, J. (2012) 'Yunlv jiegou yu yuyin de bianhua [Prosodic structure and speech sound variation]', *Report of Phonetic Research*.

Cao, J. and Zheng, Y. (2006) 'Articulatory strengthening and prosodic hierarchy', *Proceedings of the Speech Prosody*. Dreston, Germany.

Carr, P. (2013) *English Phonetics and Phonology: An Introduction*. 2nd edn. Wiley-Blackwell.

Catford, J.C. and Pisoni, D.B. (1970) 'Auditory vs. articulatory training in exotic sounds', *The Modern Language Journal*, 54(7), pp. 477-481.

Cauldwell, R. (2002) *Streaming Speech: Listening and Pronunciation for Advanced Learners of English. British/Irish version*. Birmingham: Speechinaction.

Cauldwell, R. (2013) *Phonology for Listening: Teaching the Stream of Speech*. Birmingham, UK: Speech in Action.

Celce-Murcia, M., Brinton, D. and Goodwin, J. (1996) *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. Cambridge: Cambridge University Press.

Celce-Murcia, M., Brinton, M.D., Goodwin, J. and Griner, B. (2010) *Teaching Pronunciation: A Course Book and Reference Guide*. 2nd edn. New York: Cambridge University Press.

Celik, M. VII (2001) 'Teaching English intonation to EFL/ESL students' *The Internet TESL Journal*. Available at: http://iteslj.org/Techniqus/Celik-Intonation.html.

Chafe, W. (1976) 'Givenness, contrastiveness, definitiveness, subjects, topics and points of view', in Li, C.N. (ed.) *Subject and Topic*. New York: Academic Press, pp. 25-55.

Chan, A.Y.W. and Li, D.C.-S. (2000) 'English and Cantonese Phonology in Contrast: Explaining Cantonese ESL Learners' English Pronunciation Problems', *Language Culture and Curriculum*, 13(1), pp. 67-85.

Chao, Y.R. (1930) 'A system of tone Letters', *Le Maitre Phonetique*, 45, pp. 24-27.

Chao, Y.R. (1933) 'Tone and Intonation in Chinese', *中央研究院歷史語言研究所集刊 Zhongyang Yanjiu Yuan Lishi Yuyan Yanjiusuo Jikan [Journal of Institute of History and Philology Academic Sinica]*, 4(2), pp. 121-134.

Chao, Y.R. (1968) *A Grammar of Spoken Chinese*. Berkeley, CA: University of California Press.

Chen, H. (2006) 'Intonation phrasing in Chinese EFL learners' reading speech', *Speech Prosody*. Dresden, Germany.

Chen, H. (2008a) *On Chinese EFL Learner's English Intonation Patterns (in Chinese)*. Shanghai: Shanghai Foreign Language Education Press.

Chen, H.C. (2015a) 'Acoustic analyses and intelligibility assessments of timing patterns among Chinese English learners

with different dialect backgrounds', *Journal of Psycholinguistic Research*, 44(6), pp. 749-773.

Chen, H.C. (2015b) 'Acoustic analyses and intelligibility assessments of timing patterns among Chinese English learners with different dialect backgrounds', *Journal of Psycholinguistic Research*, 44(6), pp. 749-773.

Chen, M.Y. (2000) *Tone Sandhi: Patterns Across Chinese Dialects*. Cambridge: Cambridge University Press.

Chen, P. (1999) *Modern Chinese: History and Sociolinguistics*. Cambridge: Cambridge University Press.

Chen, W. (2013) 'An empirical study on the acquisition of English rising tone by Chinese EFL learners', in *English Language Teaching*. pp. 97-102.

Chen, Y. (2008b) 'Prosodic realization of focus on verbal phrases in two Chinese dialects', *Journal of Acoustical Society of America*, 123, p. 3889.

Chen, Y., Robb, M.P., Gilbert, H.R. and Lerman, J.W. (2001) 'A study of sentence stress production in Mandarin speakers of American English', *The Journal of the Acoustical Society of America*, 109(4), pp. 1681-90.

Cheng, C.C. (1973) *A Synchronic Phonology of Mandarin Chinese, Monographs on Linguistic Analysis 4*. The Hague: Mouton.

Chun, D.M. (1998) 'Signal analysis software for teaching discourse intonation', *Language Learning & Technology*, 2(1), pp. 61-77.

Chun, D.M. (2002) *Discourse Intonation in L2: From Theory and Research to Practice*. Amsterdam/Philadelphia: John Benjamins.

Chun, D.M. (2013) 'Technology and discourse intonation', in Chapelle, C.A. (ed.) *The Encyclopedia of Applied Linguistics*. Hoboken: Blackwell Publishing, pp. 1-8.

Clifford, R. and Granoien, N. (2007) 'Applications of technology to language acquisition process: What can work and why', in Holland, M. and Fisher, P.F. (eds.) *The Path of Speech Technologies in Computer Assisted Language Learning*. New York: Routledge, pp. 25-43.

Coniam, D. (1999) 'Voice recognition software accuracy with second language speakers of English', *System*, 27(1), pp. 49-64.

Cooper, W.E., Eady, S.J. and Mueller, P.R. (1985) 'Acoustical aspects of contrastive stress in question-answer contexts', *Journal of the Acoustical Society of America*, 77, pp. 2142–2156.

Council of Europe (2001) *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge, UK: Press Syndicate of the University of Cambridge.

Couper-Kuhlen, E. and Ford, C.E. (2004) *Sound Patterns in Interaction*. Amsterdam: John Benjamins.

Cruttenden, A. (1981) 'Falls and rises: meanings and universals', *Journal of Linguistics*, 17(01), pp. 77-91.

Cruttenden, A. (1985) 'Intonation comprehension in ten-year-olds', *Journal of Child Language*, 12(03), pp. 643-661.

Cruttenden, A. (1997) *Intonation*. 2nd edn. Cambridge: Cambridge University Press.

Cruttenden, A. (2014) *Gimson's Pronunciation of English*. 8th edn. London: Routledge.

Cruz-Ferreira, M. (1987) 'Non-native interpretive strategies for intonational meaning: An experimental study', in James, A. and Leather, J. (eds.) *Sound Patterns in Second Language Acquisition*. Dordrecht: Foris, pp. 103-120.

Cruz-Ferreira, M. (1989) 'A test for non-native comprehension of intonation in English', *International Review of Applied Linguistics in Language Teaching*, 27(1), pp. 23-39.

Crystal, D. (1969) *Prosodic System and Intonation in English*. Cambridge: Cambridge University Press.

Crystal, D. (1986) 'Prosodic development', in Fletcher, P. and Garman, M. (eds.) *Language Acquisition: Studies in First Language Development.* 2nd edn. Cambridge: Cambridge University Press, pp. 174-197.

Crystal, D. (2003) *English as a Global Language*. 2nd edn. Cambridge: Cambridge University Press.

Crystal, D. (2008) *A Dictionary of Linguistics and Phonetics*. 6th edn. Malden, MA: Blackwell Publishing.

Cunnings, I. and Finlayson, I. (2015) 'Mixed effects modeling and longitudinal data analysis', in Plonsky, L. (ed.) *Advancing Quantitative Methods in Second Language Research*. New York: Routledge, pp. 159-181.

Cutler, A., Dahan, D. and van Donselaar, W. (1997) 'Prosody in the Comprehension of Spoken Language: A Literature Review', *Language and Speech*, 40(2), pp. 141-201.

Cutler, A. and Swinney, D.A. (1987) 'Prosody and the development of comprehension', *Journal of Child Language*, 14(01), pp. 145-167.

de Bot, K. (1983) 'Visual feedback of intonation I: Effectiveness and induced practice behavior', *Language and Speech*, 26(4), pp. 331-350.

Dehé, N. and Kavalova, Y. (2007) *Parentheticals*. Amsterdam: John Benjamins Publishing.

Deng, J., Holtby, A.K., Howden-Weaver, L., Nessim, L., Nicholas, B., Nickle, K., Pannekoek, C., Stephan, S. and Sun, M. (2009) 'English pronunciation research: The neglected orphan of second language acquisition studies?', *PMC Working Paper Series*, WP05-09, pp. 1-14.

Derwing, T.M. and Munro, M.J. (2001) 'What speaking rates do non-native listeners prefer?', *Applied Linguistics*, 22(3), pp. 324–337.

Derwing, T.M. and Munro, M.J. (2015) *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam, Netherland: John Benjamins Publishing Company.

Derwing, T.M., Munro, M.J. and Wiebe, G. (1997) 'Pronunciation instruction for "fossilized" learners: can it help?', *Applied Language Learning*, 8(2), pp. 217-235.

Derwing, T.M., Munro, M.J. and Wiebe, G. (1998) 'Evidence in favor of a broad framework for pronunciation instruction', *Language Learning*, 48(3), pp. 393-410.

Derwing, T.M. and Rossiter, M.J. (2002) 'ESL learners' perceptions of their pronunciation needs and strategies', *System*, 30(2), pp. 155-166.

Derwing, T.M. and Rossiter, M.J. (2003) 'The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech', *Applied Language Learning*, 13(1), pp. 1-18.

Derwing, T.M., Thomson, R.I., Foote, J.A. and Munro, M.J. (2012) 'A longitudinal study of listening perception in adult learners of English: implications for teachers', *Canadian Modern Language Review/ La Revue canadienne des langues vivantes*, 68(3), pp. 247-266.

Deterding, D. (2010) 'ELF-based pronunciation teaching in China', *Chinese Journal of Applied Linguistics*, 33(6), pp. 3-15.

Ding, H., Jokisch, O. and Hoffmann, R. (2012a) 'An audiovisual feedback system for pronunciation tutoring – Mandarin Chinese learners of German', in Esposito, A., Esposito, A., Vinciarelli, A., Hoffmann, R. and Müller, V. (eds.) *Cognitive Behavioural Systems*. Berlin Heidelberg: Springer pp. 191-197.

Ding, H., Jokisch, O. and Hoffmann, R. (2012b) 'A Phonetic Investigation of Intonational Foreign Accent in Mandarin Chinese Learners of German', *Speech Prosody*

Doughty, C. and Long, M.H. (2003) *The handbook of second language acquisition*. Malden, MA: Blackwell Pub.

Duanmu, S. (1999) 'Metrical Structure and Tone: Evidence from Mandarin and Shanghai', *Journal of East Asian Linguistics*, 8(1), pp. 1-38.

Duanmu, S. (2002) *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Duanmu, S. (2014) 'Syllable structure and stress', in Huang, C.T.J., Li, Y.H.A. and Simpson, A. (eds.) *The Handbook of Chinese Linguistics*. Hoboken, USA: John Wiley & Sons, pp. 422-442.

Eckman, F.R. (1977). Markedness and the contrastive analysis hypothesis. Language Learning, 27(2), 315-330.

Ehsani, F. and Knodt, E. (1998) 'Speech technology in computer-aided language learning: Strengths and limitations of a new CALL paradigm', *Language Learning and Technology*, 2(1), pp. 54-73.

Elliott, A.R. (1997) 'On the Teaching and Acquisition of Pronunciation within a Communicative Approach', *Hispania*, 80(1), pp. 95-108.

Engwall, O. and Bälter, O. (2007) 'Pronunciation feedback from real and virtual language teachers', *Computer Assisted Language Learning*, 20(3), pp. 235-262.

Felps, D., Bortfeld, H. and Gutierrez-Osuna, R. (2009) 'Foreign accent conversion in computer assisted pronunciation training', *Speech Communication*, 51(10), pp. 920-932.

Field, A., Miles, J. and Field, Z. (2012) *Discovering Statistics using R*. London: Sage.

Flege, J.E. (1995) 'Second language speech learning: Theory, findings, and problems', in Strange, W. (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore: York Press, pp. 233-277.

Flege, J.E., Schirru, C. and MacKay, I.R.A. (2003) 'Interaction between the native and second language phonetic subsystems', *Speech Communication*, 40(4), pp. 467-491.

Gananathan, R.Y., Yin, J. and Mok, P. (2015) 'Interlanguage influence in cues of narrow focus: A study of Hong Kong English', *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*. Glasgow, Scotland.

Gao, W. (2015) *Exploring a Functional Prosodic Approach in the Teaching of English Intonation in Chinese Context--An Experimental Study*. Doctoral thesis. Shanghai International Studies University.

Geranpayeh, A. (2003) 'A quick review of the English Quick Placement Test', *Research notes*, 12, pp. 8-10.

Gili Fivela, B. (2012) 'Testing the perception of L2 intonation', in Busa, M.G. and Stella, A. (eds.) *Methodological Perspectives on Second Language Prosody. Papers from ML2P 2012*. Padova: CLEUP, pp. 17-30.

Godwin-Jones, R. (2009) 'Emerging technologies speech tools and technologies', *Language Learning & Technology*, 13(3), pp. 4-11.

Gorjian, B., Hayati, A. and Pourkhoni, P. (2013) 'Using Praat software in teaching prosodic features to EFL learners', *Procedia - Social and Behavioral Sciences*, 84(0), pp. 34-40.

Grabe, E., Kochanski, G. and Coleman, J. (2005) 'The intonation of native accent varieties in the British Isles: Potential for miscommunication?', in Dziubalska-Kołaczyk, K. and Przedlacka, J. (eds.) *English Pronunciation Models: A Changing Scene*. Bern: Peter Lang, pp. 311-337.

Greif, M. (2010) 'Contrastive focus in mandarin Chinese', *Proceedings of Speech Prosody*.

Grosser, W. (1997) 'On the acquisition of tonal and accentual features of English by Austrian learners', in James, A. and Leather, J. (eds.) *Second Language Speech: Structure and Process.* Berlin: Mouton de Gruyter, pp. 211-228.

Gu, Y., Jin, Y., Zou, S., Yang, Z., Wu, Z., He, L., Zhang, W., Huang, S., Jia, G., Tang, Q., Wang, S. and Lin, H. (2013) *Standards for Oral Proficiency in English*. Beijing, China: Renmin University of China Press.

Gussenhoven, C. (2002) 'Intonation and interpretation: phonetics and phonology', *The Proceedings of Speech Prosody*. aix-en-Provence, France. pp. 47-57.

Gut, U. (2009) *Non-native Speech: A Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German*. Frankfurt am Main; Oxford: Peter Lang.

Gut, U., Pillai, S. and Don, Z.M. (2013) 'The prosodic marking of information status in Malaysian English', *World Englishes*, 32(2), pp. 185-197.

Gut, U., Trouvain, J. and Barry, W.J. (2007) 'Bridging research on phonetic descriptions with knowledge from teaching practice-The case of prosody in non-native speech', in Trouvain, J. and Gut, U. (eds.) *Non-Native Prosody : Phonetic Description and Teaching Practice*. Gruyter, Berlin: De Gruyter, pp. 3-24.

Halliday, M.A.K. (1967a) *Intonation and Grammar in British English*. The Hague: Mouton.

Halliday, M.A.K. (1967b) 'Notes on transitivity and theme in English, Part 2', *Journal of Linguistics*, 3(2), pp. 199-244.

Halliday, M.A.K. (1970) *A Course in Spoken English: Intonation*. London: Oxford University Press.

Halliday, M.A.K. (2004) *An Introduction to Functional Grammar*. 3rd edn. London: Arnold Ltd.

Hardison, D.M. (2004) 'Generalization of computer-assisted prosody training: Quantitative and qualitative findings', *Language Learning & Technology*, 8(1), pp. 34-52.

Hardison, D.M. (2005) 'Contextualised computer-based L2 prosody training: Evaluating the effects of discourse context and video input', *CALICO Journal*, 22(2), pp. 175-190.

Hawkins, R. and Casillas, G. (2008) 'Explaining frequency of verb morphology in early L2 speech', *Lingua*, 118(4), pp. 595-612.

Hazan, V., Sennema, A., Iba, M. and Faulkner, A. (2005) 'Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English', *Speech Communication*, 47(3), pp. 360-378.

He, X., Hanssen, J., Van Heuven, V.J. and Gussenhoven, C. (2011) 'Phonetic implementation must be learnt: Native versus Chinese realisation of focal accent in Dutch', *Proceesings of the 17th International Congress of Phonetic Sciences*. Hong Kong. pp. 843-846.

He, X., Hanssen, J., van Heuven, V.J. and Gussenhoven, C. (2012a) 'Mandarin-accented Fall, Rise and Fall-Rise F0 contours in Dutch', *Speech Prosody, 6th International Conference*. Shanghai, China.

He, X., van Heuven, V.J. and Gussenhoven, C. (2010) 'Choosing the optimal pitch accent location in Dutch by Chinese learners and native listeners', *Proc. 6th New Sounds*.

He, X., van Heuven, V.J. and Gussenhoven, C. (2012b) 'The selection of intonation contours by Chinese L2 speakers of Dutch: Orthographic closure vs. prosodic knowledge', *Second Language Research*, 28(3), pp. 283-318.

Hermes, D.J. (1998) 'Auditory and visual similarity of pitch contours', *Journal of Speech, Language, and Hearing Research*, 41(1), pp. 63-72.

Hewings, M. (1995) 'Tone choice in the English intonation of non-native speakers', *International Review of Applied Linguistics*, 33(3), pp. 251-284.

Hincks, R. (2003) 'Speech technologies for pronunciation feedback and evaluation', *ReCALL*, 15(01), pp. 3-20.

Hincks, R. and Edlund, J. (2009) 'Promoting increased pitch variation in oral presentations with transient visual feedback', *Language Learning & Technology*, 13(3), pp. 32-50.

Ho, A.T. (1977) 'Intonation variation in a Mandarin sentence for three expressions: interrogative, exclamatory and declarative', *Phonetica*, 34(6), pp. 446-457.

Jaeger, T.F. (2008) 'Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models', *Journal of memory and language*, 59(4), pp. 434-446.

Jenkins, J. (2000) *The Phonology of English as an International Language*. Oxford: Oxford University Press.

Jenkins, J. (2004) 'Research in teaching pronunciation and intonation', *Annual Review of Applied Linguistics*, 24, pp. 109-125.

Jenkins, J. (2007) *English as a Lingua Franca: Attitude and Identity*. Oxford, UK: Oxford University Press.

Ji, X., Wang, X. and Li, A. (2009) 'Intonation patterns of yes-no questions for Chinese EFL learners', *Speech Database and Assessments, 2009 Oriental COCOSDA International Conference*. 10-12 Aug. 2009. pp. 88-93.

Jia, G. (2017) 'Daxue Yingyu Jiaoxue Zhinan yu gaoxiao daxue yingyu jiaoxue gaige [Guidelines on College English Teaching and the reforms of college English teaching]', *Dangdai Waiyu Yanjiu [Contemporary Foreign Languages Studies]*, 432(6), pp. 62-65.

Jia, Y., Li, A. and Xiong, Z. (2010) 'A phonetic and phonological analysis of dual and multiple focuses in Standard Chinese', *Report of Phonetic Research*, pp. 48-51.

Jia, Y., Xiong, Z. and Li, A. (2008) 'Putonghua jiaodian zhongyin dui yuju yingao de zuoyong [The effects of focal accents on sentential F0 in Mandarin]', *Zhongguo Yuyin Xuebao [Chinese Journal of Phonetics]*, 1, pp. 118-124.

Jiang, P. and Chen, A. (2011) 'Representation of Mandarin intonations: boundary tone revisited', *Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-23)*. University of Oregon, Eugene. pp. 97-109.

Johnson, K. (1997) 'Speech perception without speaker normalization', in Johnson, K. and Mullennix, J. (eds.) *Talker variability in speech processing*. San Diego: Academic Press.

Jun, S.-A. and Bishop, J. (2015) 'Priming implicit prosody: prosodic boundaries and individual differences', *Language and Speech*, 58(4), pp. 459-473.

Kingdon, R. (1958) *The Ground Work of English Intonation*. London: Longmans.

Kratochvil, P. (1998) 'Intonation in Beijing Chinese', in Hirst, D. and Di Cristo, A. (eds.) *Intonation Systems: A Survey of Twenty Languages*. Oxford: Oxford University Press, pp. 417-432.

Kurt, B.G., Medlin, J. and Tessarolo, A. (2014) 'The Perception of Prosodically Ambiguous Intonation Patterns by L2 English Learners and the Effects of Instruction', *Concordia Working Papers in Applied Linguistics*, 5, pp. 353-372.

Ladd, D.R. (1996) *Intonational Phonology*. Cambridge: Cambridge University Press.

Ladd, D.R. (2008) *Intonational Phonology*. 2nd edn. Cambridge: Cambridge University Press.

Ladefoged, P. (2001) *A Course in Phonetics*. Fort Worth, TX: Harcourt College Publishers.

Lado, R. (1957) *Linguistics across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor, MI: University of Michigan Press.

Le, H.T. and Brook, J. (2011) 'Using Praat to teach intonation to ESL students', *Hawaii Pacific University TESOL Working Paper Series*, 9(1, 2), pp. 2-15.

Lee, Y.-C. and Xu, Y. (2010) 'Phonetic realization of contrastive focus in Korean', *Speech Prosody*. Chicago.

Lenth, R.V. (2016) 'Least-Squares Means: The R Package lsmeans', *Journal of Statistical Software*, 69(1), pp. 1-33.

Lepage, A. and Busà, M.G. (2014) 'Intelligibility of English L2: The Effects of Incorrect Word Stress Placement and Incorrect Vowel Reduction in the Speech of French and Italian Learners of English', *Concordia Working Papers in Applied Linguistics*, 5, pp. 387-400.

Levis, J.M. (1999) 'Intonation in Theory and Practice, Revisited*', *TESOL Quarterly*, 33(1), pp. 37-63.

Levis, J.M. (2005) 'Comparing apples and oranges? Pedagogical approaches to intonation', in Dziubalska-Kołaczyk, K. and Przedlacka, J. (eds.) *English Pronunciation Models: A Changing Scene*. Bern: Peter Lang, pp. 339-366.

Levis, J.M. (2007) 'Computer technology in teaching and researching pronunciation', *Annual Review of Applied Linguistics*, 27, pp. 184-202.

Levis, J.M. and Wichmann, A. (2015) 'English Intonation–Form and Meaning', in Reed, M. and Levis, J.M. (eds.) *The Handbook of English pronunciation*. Hoboken, NJ: Wiley Blackwell, pp. 139-155.

Levy, M. (1997) *Computer-assisted Language Learning: Context and Conceptualization*. Oxford: Oxford University Press.

Li, A. and Post, B. (2014) 'L2 acquisition of prosodic properties of speech rhythm', *Studies in Second Language Acquisition*, 36(02), pp. 223-255.

Li, A., Yin, Z. and Zu, Y. (2006) 'A rhythmic analysis on Chinese EFL speech', *Speech Prosody*. Dresden, Germany.

Li, W. (2012) 'Jiaodian de biaoji shouduan ji renzhi jiagong [The markers of focus and its cognitive processing]', *Chinese Journal of Phonetics*, 3, pp. 61-69.

Liberman, M. and Prince, A. (1977) 'On stress and linguistic rhythm', *Linguistic Inquiry*, 8(2), pp. 249-336.

Lightbown, P.M. and Spada, N. (2006) *How Languages are Learned*. 3rd edn. Oxford, UK: Oxford University Press.

Lin, M. (2004a) 'Boundary Tone of Chinese Intonation and its Pitch (F0) Pattern', *Report of Phonetic Research*, pp. 9-18.

Lin, M. (2004b) 'Hanyu yudiao yu shengdiao [Chinese intonation and tone]', *Yuyan Wenzi Yingyong [Applied Linguistics]*, 3, pp. 57-67.

Lin, M. (2006) 'Interrogative vs. declarative and boundary tone', *Zhongguo Yuwen*, 4, pp. 364-376.

Lin, M. (2008) 'Chao Yuan-Ren hanyu yudiao sixiang yu yiwen yudiao he chengshu yudiao: Jiantan yudiao jiaoxue [Review of Chao Yuan-Ren's intonation system with explanation of interrogative and declarative intonation: An implication for intonation teaching]', *Zhongguo Yuyin Xuebao [Chinese Journal of Phonetics]*, 1, pp. 64-70.

Lin, M. and Li, Z. (2011) 'Focus and boundary in Chinese intonation', *Proceedings of International Congress of Phonetic Sciences*. Hong Kong, China. pp. 1246-1249.

Lin, Y.-H. (2007) *The Sounds of Chinese*. Cambridge: Cambridge University Press.

Liu, C. and Rodriguez, A. (2012) 'Categorical perception of intonation contrasts: effects of listeners' language background', *The Journal of the Acoustical Society of America*, 131(6), pp. EL427-33.

Liu, F. and Xu, Y. (2005) 'Parallel Encoding of Focus and Interrogative Meaning in Mandarin Intonation', *Phonetica*, 62(2-4), pp. 70-87.

Liu, F. and Xu, Y. (2007) 'Question intonation as affected by word stress and focus in English', *16th International Congress of Phonetic Science*. Saarbrucken, Germany. pp. 1189-1192.

Liu, F., Xu, Y., Prom-on, S. and Yu, A.C.L. (2013) 'Morpheme-like prosodic functions: Evidence from acoustic analysis and computational modeling', *Journal of Speech Sciences*, 3(1), pp. 85-140.

Liu, X. and Chen, X. (2016) 'The acquisition of English pitch accents by Mandarin Chinese speakers as affected by boundary tones', *Speech Prosody*. Boston, USA. pp. 956-960.

Local, J. and Walker, G. (2004) 'Abrupt joins as a resource for he production of multi-unit, multi-action turns', *Journal of Pragmatics*, 36(8), pp. 1375-1403.

Lord, G. (2010). The combined effects of immersion and instruction on second language pronunciation. *Foreign language annals*, *43*(3), 488-503.

Lu, X. (2010) 'Daxue yingyu zhuanye yuyin jiaoxue de xianzhuang diaocha yu wenti fenxi [An investigation of the current status of the teaching of Enlgish pronunciation in China's higher education system]', in Chen, H. and Yang, J. (eds.) *Zhonguo Eryu Yuyin Xide Yanjiu De Xianzhuang Ji Fazhan Qushi [Second Language Phonological Studies in China]*. Beijing: Foreign Language Teaching and Research Press, pp. 541-560.

Luo, B. (2016) 'Evaluating a computer-assisted pronunciation training (CAPT) technique for efficient classroom instruction', *Computer Assisted Language Learning*, 29(3), pp. 451-476.

MacDonald, D. (2011) 'Second language acquisition of English question in intonation by Koreans', *Proceedings of the 2011 annual Conference of the Canadian Linguistic Association*. University of New Brunswick, Fredericton. pp. 1-15.

Mair, P., Schoenbrodt, F. and Wilcox, R. (2016) *Wilcox robust estimation and testing*.

Manning, C. (2007) *Generalized Linear Mixed Models (illustrated with R on Bresnan et al.'s datives data)*. Available at: http://www-nlp.stanford.edu/manning/courses/ling289/GLMM.pdf.

Martin, P. (2004) 'WinPitch LTL II, a multimodal software', *Proceedings of InSTIL/ICALL*. Venice, Italy.

Massaro, D.W. (1998) 'Categorical perception: Important phenomenon or lasting myth?', *5th International Conference on Spoken Language Processing*. Sydney, Australia. pp. 2275-2279.

Mennen, I. (2004) 'Bi-directional interference in the intonation of Dutch speakers of Greek', *Journal of Phonetics*, 32(4), pp. 543-563.

Mennen, I. (2006) 'Phonetic and phonological influences in non-native intonation: an overview for language teachers', *Working Papers Queen Margaret University College*, 9.

Mennen, I. (2007) 'Phonological and phonetic influences in non-native intonation', in Trouvain, J. and Gut, U. (eds.) *Nonnative Prosody: Phonetic Description and Teaching Practice*. Berlin: Mouton de Gruyter, pp. 53-76.

Mennen, I. (2015) 'Beyond segments: Towards a L2 intonation learning theory', in Delais-Roussarie, E., Avanzi, M. and Herment, S. (eds.) *Prosody and Language in Contact: L2 Acquisition, Attrition and Languages in Multilingual Situations*. Berlin: Springer, pp. 171-188.

Mennen, I., Chen, A. and Karlsson, F. (2010) 'Characterising the internal structure of learner intonation and its development over time', *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds*. Poznan: Adam Mickiewicz University. pp. 319-324.

Mennen, I. and de Leeuw, E. (2014) 'Beyond Segments: Prosody in SLA', *Studies in Second Language Acquisition*, 36(02), pp. 183-194.

Mennen, I., F. Schaeffler and Dickie., C. (2014) 'Second language acquisition of pitch range in German learners of English', *Studies in Second Language Acquisition*, 36(2), pp. 303-329.

Mimutsu (2000) *An essay on incorrect English pronunciation of text reading by Japanese college students*. Available at: http://www.lingua.tsukuba.ac.jp/ippan/JGL/2000/2000-Mimatsu.pdf.

Mitrofanova, Y. (2012) 'Raising EFL students' awareness of English intonation functioning', *Language Awareness*, 21(3), pp. 279-291.

Mok, P., Yin, Y., Setter, J. and Nayan, N.M. (2016) 'Assessing knowledge of English intonation patterns by L2 speakers', *Speech Prosody*. Boston, MA. pp. 543-547.

Morley, J. (1991) 'The pronunciation component in teaching English to speakers of other languages', *TESOL Quarterly*, 25(3), pp. 481-520.

Munro, M. J. (2003). A primer on accent discrimination in the Canadian context. *TESL Canada Journal*, 38-51.

Munro, M.J. and Derwing, T.M. (1995) 'Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners', *Language Learning*, 45(1), pp. 73-97.

Munro, M.J. and Derwing, T.M. (1998) 'The effects of speaking rate on listener evaluations of native and foreign-accented speech', *Language Learning*, 48(2), pp. 159-182.

Munro, M.J. and Derwing, T.M. (2001) 'Modelling perceptions of the comprehensibility and accentedness of L2 speech: the role of speaking rate', *Studies in Second Language Acquisition*, 23(4), pp. 451-468.

Munro, M.J. and Derwing, T.M. (2015) 'A prospectus for pronunciation research in the 21st century: A point of view', *Journal of Second Language Pronunciation*, 1(1), pp. 11-42.

Murphy, J.M. and Baker, A.A. (2015) 'History of ESL pronunciation teaching', in Reed, M. and Levis, J.M. (eds.) *The Handbook of English Pronunciation*. pp. 36-65.

Murray, G.L. (1999) 'Autonomy and language learning in a simulated environment', *System*, 27(3), pp. 295-308.

Neri, A., Cucchiarini, C., Strik, H. and Boves, L. (2002) 'The pedagogy-technology interface in computer assisted pronunciation training', *Computer Assisted Language Learning*, 15(5), pp. 441-467.

Nicolaidis, K., Papanikolaou, G., Kainada, E. and Avdelidis, K. (2015) 'SpeakGreek: An online speech training tool for L2 pedagogy and clinical intervention', *Proceedings of the International Congress of Phonetic Sciences*. Glasgow, Scotland.

Nilsen, D.L.F. and Nilsen, A.P. (1971) *Pronunciation Contrasts in English*. New York, NY: Simon and Schuster.

Nolan, F. (2006) 'Intonation', in Aarts, B. and McMahon, A. (eds.) *The Handbook of English Linguistics*. Oxford: Blackwell Publishing, pp. 433-457.

O'Connor, J.D. and Arnold, G.F. (1961) *Intonation of Colloquial English*. London: Longman.

O'Brien, M. and Gut, U. (2010) 'Phonological and phonetic realisation of different types of focus in L2 speech', *Achievements and perspectives in the acquisition of second language speech: New sounds 2010*, pp. 205-215.

Olson, D.J. (2014) 'Benefits of visual feedback on segmental production in the L2 classroom', *Language Learning & Technology*, 18(3), pp. 173-192.

Ortega-Llebaria, M. and Colantoni, L. (2014) 'L2 English intonation: Relations between form-meaning associations, access to meaning, and L1 transfer', *Studies in Second Language Acquisition*, 36(02), pp. 331-353.

Palmer, H.E. (1922) *English Intonation with Systematic Exercises*. Cambridge: Heffer.

Pan, H., Huang, W. and Huang, Y. (2005) 'Production and perception of new, given, and contrastive information in Taiwan Mandarin', *Proceedings of ISCA Workshop on Plasticity in Speech Perception*. London, UK.

Pennington, M.C. and Richards, J. (1986) 'Pronunciation revisited', *TESOL Quarterly*, 20(2), pp. 207-225.

Pierrehumbert, J. (1980) *The phonetics and phonology of English intonation*. Unpublished doctoral dissertation thesis. MIT.

Pierrehumbert, J. (2017) 'Comparing PENTA to autosegmental-metrical phonology. In press', in Barnes, J. and Shattuck-Hufnagel, S. (eds.) *Prosodic Theory and Practice*. MIT Press.

Pierrehumbert, J. and Beckman, M. (1988) *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Pierrehumbert, J. and Hirschberg, J. (1990) 'The meaning of intonational contours in the interpretation of discourse', in Cohen, P.R., Morgan, J. and Pollack, M.E. (eds.) *Intentions in Communication*. Cambridge MA: MIT Press, pp. 271-311.

Pike, K.L. (1945) *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Prator, C. and Robinett, B.W. (1985) *Manual of American English Pronunciation*. 4th edn. New York: Holt, Rinehart & Winston.

Prom-on, S., Xu, Y. and Thipakorn, B. (2009) 'Modeling tone and intonation in Mandarin and English as a process of target approximation', *Journal of Acoustical Society of America*, 125(1), pp. 405-424.

Puga, K., Fuchs, R., Setter, J. and Mok, P. (2017) 'The perception of English intonation patterns by German L2 speakers of English', *INTERSPEECH*. Stockholm, Sweden. pp. 3241-3245.

Quené, H. and van den Bergh, H. (2008) 'Examples of mixed-effects modeling with crossed random effects and with binomial data', *Journal of Memory and Language*, 59(4), pp. 413-425.

Quick Placement Test (2001) *Paper and pen test*. Available at: http://www.bz.nuernberg.de/fileadmin/bilder/Sprachen/Englisch/pdf/Einstufung_QPT.pdf.

R Core Team (2014) *R: A language and environment for statistical computing* [Computer program]. R Foundation for Statistical Computing, Vienna, Austria.

Ramírez Verdugo, D. (2003) 'Non-native interlanguage intonation systems: A study based on a computerized corpus of Spanish learners of English', *ICAME*, 26, pp. 115-132.

Ramírez Verdugo, D. (2006) 'A study of intonation awareness and learning in non-native speakers of English', *Language Awareness*, 15(3), pp. 141-159.

Rasier, L. and Hiligsmann, P. (2007) 'Prosodic transfer from L1 to L2. Theoretical and methodological issues', *Nouveaux Cahiers de Linguistique Française*, 28, pp. 41-66.

Richards, J.C. and Schmidt, R. (2002) *Longman Dictionary of Language Teaching & Applied Linguistics*. 3rd edn. Harlow: Longman an imprint of pearson.

Roach, P. (2000) *English Phonetics and Phonology*. 3rd edn. Cambridge: Cambridge university Press.

Roach, P. (2009) *English Phonetics and Phonology: A Practical Course*. 4th edn. Cambridge: Cambridge University Press.

Rogerson-Revell, P. (2011). *English phonology and pronunciation teaching*. New York: Bloomsbury Publishing.

Rosenberg, A., Hirschberg, J. and Manis, K. (2010) 'Perception of English prominence by native Mandarin Chinese speakers', *Speech Prosody*. Chicago, Illinois. Available at: http://speechprosody2010.illinois.edu/papers/100982.pdf.

Saalfeld, A. (2011) 'Acquisition of L2 phonology in advanced learners: Does instruction make a difference?', *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference*. Iowa State University. pp. 144-152.

Saito, K. (2012) 'Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies', *TESOL Quarterly*, 46(4), pp. 842-854.

Sánchez-Alvarado, C. and Armstrong, M.E. (2016) 'Pitch scaling and the perception of contrastive focus in L1 and L2 Spanish', *Speech Prosody*. Boston, USA. pp. 592-596.

Schack, K. (2000) 'Comparison of intonation patterns in Mandarin and English for a particular speaker', *University of Rochester Working Papers in the Language Science*, Spring 2000(1), pp. 24-55 [Online].

Setter, J. (2006) 'Speech Rhythm in World Englishes: The Case of Hong Kong', *TESOL Quarterly*, 40(4), pp. 763-782.

Setter, J. and Jenkins, J. (2005) 'State-of-the-Art Review Article: Pronunciation', *Language Teaching*, 38(01), pp. 1-17.

Shao, P., Jia, Y. and Li, A. (2011) 'Intonation patterns of exclamations of Chinese EFL learners from Zhenjiang', *International Conference on Speech Database and Assessments (Oriental COCOSDA)*. Hsinchu, Taiwan. IEEE, pp. 74-78.

Shen, X.-N. (1989) *The Prosody of Mandarin Chinese*. CA: University of California Press.

Shi, B. (2010) 'Yingyu zhuanye yuyinke jiaoxue de jige zhuyao wenti [A review of the pronunciation teaching for English majors]', in Chen, H. and Yang, J. (eds.) *Zhonguo Eryu Yuyin Xide Yanjiu De Xianzhuang Ji Fazhan Qushi [Second Language Phonological Studies in China]*. Beijing, China: Foreign Language Teaching and Research Press, pp. 59-65.

Shih, C. (2001) *Generation and normalization of tonal variations*.

Silverman, K.E.A., Beckman, M.E., Pitrelli, J.F., Ostendorf, M., Wightman, C.W., Price, P., Pierrehumbert, J. and Hirschberg, J. (1992) 'ToBI: a standard for labeling English prosody', *In Proceedings of The 1992 International Conference on Spoken Language Processing*. Banff. pp. 867-870.

Singmann, H., Bolker, B., Westfall, J. and Aust, F. (2016) *afex: Analysis of Factorial Experiments. R package version 0.16-1.* Available at: https://CRAN.R-project.org/package=afex.

Sluijter, A. and van Heuven, V. (1996) 'Spectral balance as an acoustic correlate of linguistic stress', *Journal of the Acoustical Society of America*, 100(4), pp. 2471-2485.

So, C.K. and Best, C.T. (2008) 'Do English speakers assimilate Mandarin tones to English prosodic categories?', *Proceedings of Interspeech 2008 incorporating SST 2008*. Brisbane.

So, C.K. and Best, C.T. (2010) 'Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences', *Language and speech*, 53(2), pp. 273-293.

So, C.K. and Best, C.T. (2011) 'Categorizing Mandarin tones into listeners' native prosodic categories: The role of phonetic properties', *Poznań Studies in Contemporary Linguistics*, 47(1), pp. 133-145.

So, C.K. and Best, C.T. (2014) 'Phonetic Influences on English and French Listeners' Assimilation of Mandarin Tones to Native Prosodic Categories', *Studies in Second Language Acquisition*, 36(2), pp. 195-221.

Song, Y. and Lan, D. (2010) 'Yudiao yishi zai yingyu jiaoji nengli peiyang zhong de weizhi [The role of intonation awareness in English communicative skill]', in Chen, H. and Yang, J. (eds.) *Zhonguo Eryu Yuyin Xide Yanjiu De Xianzhuang Ji Fazhan Qushi [Second Language Phonological Studies in China]*. Beijing: Foreign Language Teaching and Research Press.

Tajima, K., Port, R. and Dalby, J. (1997) 'Effects of temporal correction on intelligibility of foreign-accented English', *Journal of Phonetics*, 25(1), pp. 1-24.

Tanner, M.W. and Landon, M.M. (2009) 'The effects of computer-assisted pronunciation reading on ESL learners' use of pausing, stress, intonation, and overall comprehensibility', *Language Learning & Technology*, 13(3), pp. 51-65.

Tench, P. (1996) *The Intonation Systems of English*. London: Cassell.

The Ministry of Education in China (2007) *College English Curriculum Requirements* [Computer program]. Shanghai Foreign Language Education Press.

Thomson, R.I. and Derwing, T.M. (2014) 'The effectiveness of L2 pronunciation instruction: A narrative review', *Applied Linguistics*, 36(3), pp. 326-344.

Tian, Z. (2010) 'Weirao ju zhongyin he jiezou de yingyu yuyin jiaoxue [English pronunciaiton teaching focused on sentence stress and rhythm]', in Chen, H. and Yang, J. (eds.) *Zhonguo Eryu Yuyin Xide Yanjiu De Xianzhuang Ji Fazhan Qushi [Second Language Phonological Studies in China]*. Beijing: Foreign Language Teaching and Research Press, pp. 97-104.

Timkova, R. (2001) 'Intonation of English in the process of second language acquisition'. pp. 55-58. Available at: http://www.phon.ucl.ac.uk/home/johnm/ptlc2001/pdf/timkova.pdf.

Toivanen, J. (2003) 'Tone choice in the English intonation of proficient non-native speakers', *Proceedings of Fonetik*. Umea/Lovanger, Swede. pp. 165-168.

Trager, G. and Smith, H. (1951) *An Outline of English Structure*. Norman, OK: Battenburg.

Trofimovich, P. and Baker, W. (2006) 'Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech', *Studies in Second Language Acquisition*, 28(01), pp. 1-30.

Trouvain, J. and Gut, U. (2007) *Non-Native Prosody : Phonetic Description and Teaching Practice*. De Gruyter. Available at: http://NCL.eblib.com/patron/FullRecord.aspx?p=364707.

Trouvain, J., Jugler, J. and Laprie, Y. (2016) 'On the assessment of computer-assisted pronunciation training tools', *Proceedings of 26th Konferenz Elektronische Sprachsignalverarbeitung (ESSV'16)*. Leipzig. pp. 175-182.

Usó-Juan, E. and Martínez-Flor, A. (2015) 'Assessing EFL Learners' Performance of the Conventional Expressions of Complaining and Apologising', *Procedia - Social and Behavioral Sciences*, 173, pp. 53-60.

Wang, A. (2003) *Research on the Pitch Downtrend of Intonation in Putonghua*. Beijing University.

Wang, B. and Xu, Y. (2011) 'Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese', *Journal of Phonetics*, 39(4), pp. 595-611.

Wang, B., Xu, Y. and Ding, Q. (2018) 'Interactive prosodic marking of focus, boundary and newness in Mandarin', *Phonetica*, 75(1), pp. 24-56.

Wang, G. (2005) *English Pronunciation & Intonation for Communication: A Course for Chinese EFL learners* 2nd edn. Beijing: Higher Education Press

Wang, G. (2010) 'Guanyu yingyu yuyin kecheng jianshe de jidian sikao [A few constructive comments on the development of English pronunciation curriculum]', in Chen, H. and Yang, J. (eds.) *Zhonguo Eryu Yuyin Xide Yanjiu De Xianzhuang Ji Fazhan Qushi [Second Language Phonological Studies in China]*. Beijing, China: Foreign Language Teaching and Research Press, pp. 15-31.

Wang, G. (2011) *English Pronunciation and Intonation: Practice* 2nd edn. Beijing: Higher Education Press.

Wang, J. (2013) 'Guowai chuantong yunlv yanjiu de zuixin jinzhan ji sikao [Latest overseas developments in the study of prosody and some reflections]', *Journal of Foreign Languages*, 36(6), pp. 18-26.

Wang, S. (2016) 'Daxue Yingyu Jiaoxue Zhinan yaodian jiedu [Review of Guidelines on College English Teaching]', *Waiyujie [Foreign Language World]*, 174(3), pp. 2-10.

Wang, S. and Wang, H. (2011) 'Woguo gaoxiao daxue yingyu jiaoxue xianzhuang diaocha ji daxue yingyu jiaoxue gaige yu fazhan fangxiang [On the state of college English teaching in China and its future development]', *Zhongguo waiyu [Foreign Languages in China]*, 8(5), pp. 4-11, 17.

Wang, W.S.-Y. (1967) 'Phonological features of tone', *IJAL*, 33(2), pp. 93-105.

Wang, X., Jia, Y. and Li, A. (2011a) 'A comparative study on accentuation implementation of Chinese EFL learners vs. American native speakers', *Speech Database and Assessments (Oriental COCOSDA)*. 26-28 Oct. 2011. pp. 62-67.

Wang, X., Li, A. and Ji, X. (2010a) 'Perception and production of prominence distribution patterns of Chinese EFL learners', *Speech Prosody 2010-Fifth International Conference*. Chicago, Illinois.

Wang, X., Li, A., Sun, J. and Mai, Y. (2010b) 'Prosodic analysis on English mild imperatives of Chinese EFL learners', *Report of Phonetic Research*, pp. 95-98 [Online]. Available at: http://paslab.phonetics.org.cn/wp-content/files/research_report/2010/2010_16.pdf.

Wang, Y. and Chu, M. (2008) 'On lexical stress of Mandarin Chinese [Guanyu putonghua cizhongyin de ruogan wenti]', *Chinese Journal of Phonetics [Zhongguo Yuyin Xuebao]*, 1, pp. 141-147.

Wang, Y., Liang, B., Wang, M., Hou, X., Zhang, X., Wang, Y., Han, Y. and Shi, X. (2011b) *English Pronunciation for College Students: A Fundamental and Advanced Course* Beijing: Higher Education Press.

Wang, Z. and Feng, S. (2006) 'Tonal contrast and disyllabic stress patterns in Beijing Mandarin', *Yuyan Kexue [Linguistic Science]*, 5(1), pp. 3-22.

Wells, J.C. (2006) *English Intonation: An Introduction*. Cambridge: Cambridge University Press.

Wennerstorm, A. (1994) 'Intonational meaning in English discourse: A study of non-native speakers', *Applied Linguistics*, 15(4), pp. 399-420.

Wichmann, A. (2000) *Intonation in Text and Discours*. London: Longman.

Wilson, I. (2008) 'Using Praat and Moodle for teaching segmental and suprasegmental pronunciation', *In Proceedings of the 3rd International WorldCALL Conference: Using Technologies for Language Learning*. Fukuoka, Japan.

Xu, Y. (1999) 'Effects of tone and focus on the formation and alignment of F0 contours', *Journal of Phonetics*, 27, pp. 55-105.

Xu, Y. (2004a) 'Transmitting tone and intonation simultaneously—The Parallel Encoding and Target Approximation (PENTA) model', *Proceedings of International Symposium on Tonal Aspects og Languages: With Emphasis on Tone Languages*. Beijing, The Institute of Linguistics in Chinese Academy of Social Sciences. pp. 215-220.

Xu, Y. (2004b) 'Understanding tone from the perspective of production and perception', *Language and Linguistics*, 5, pp. 757-797.

Xu, Y. (2005) 'Speech melody as articulatorily implemented communicatice functions', *Speech Communication*, 46, pp. 220-251.

Xu, Y. (2012) 'Function vs. Form in Speech Prosody – Lessons from Experimental Research and Potential Implications for Teaching', in Romero-Trillo, J. (ed.) *Pragmatics and Prosody in English Language Teaching*. Springer Netherlands, pp. 61-76.

Xu, Y. (2017) 'Intonation', in Behr, W., Yue, G.G., ZevHandel, C.T., Huang, J. and Myers, J. (eds.) *Encyclopedia of Chinese Language and Linguistics*. Boston: Brill, pp. 458-466.

Xu, Y., Chen, S.W. and Wang, B. (2012) 'Prosodic focus with and without post-focus compression: A typological divide within the same language family?', *The Linguistic Review*, 29(1), pp. 131-147.

Xu, Y. and Wang, M. (2009) 'Organizing syllables into groups—Evidence from F0 and duration patterns in Mandarin', *Journal of Phonetics*, 37(4), pp. 502-520.

Xu, Y. and Wang, Q.E. (1997) 'What can tone studies tell us about intonation?', *Intonation: Theory, Models and Applications*. Athens, Greece.

Xu, Y. and Xu, C.X. (2005) 'Phonetic realization of focus in English declarative intonation', *Journal of Phonetics*, 33(2), pp. 159-197.

Yang, J. (2010) 'Fanfu linting mofang he zhongguo yingyu xuesheng zhongyin moshi xide yanjiu ["Listen and Repeat" and Chinese learners' acquisition of English stress patterns]', in Chen, H. and Yang, J. (eds.) *Zhonguo Eryu Yuyin Xide Yanjiu De Xianzhuang Ji Fazhan Qushi [Second Language Phonological Studies in China]*. Beijing: Foreign Language Teaching and Research Press, pp. 561-578.

Yang, D., Li, N., Liu, Y., Zhang, S., Yang, D. and Qin, X. (2013) *A practical Course for English Pronunciation Skills* Beijing: China Book Press.

Yip, M. (2002) *Tone*. Cambridge: Cambridge University Press.

Yuan, C., Gonzalez-Fuente, S., Baills, F. and Prieto, P. (2017) 'Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers', *Studies in Second Language Acquisition*, pp. 1-28.

Zeng, Y. (2017) *An analysis of English intonation features of Chinese junior high school English teachers: A case study of the teacher training school of Qingpu district in Shanghai*. Master's thesis. Shanghai Normal University.

Zhang, C. (2015) 'The impact of audio-visual media on English phonological acquisition and defossilization by learners in China', *Newcastle and Northumbria Working Papers in Linguistics*, 21(2), pp. 81-97.

Zhang, F. (2003) *British English Phonetics and Phonology*. 3rd edn. Chengdu, Sichuan: Sichuan University Press.

Zhang, G. (2006) *A Handbook of English Pronunciation Exercises for First-Year College Students*. 2nd edn. Beijing: Foreign Language Teaching and Research Press

Zhang, G. (2008) *British English Intonation: Studies in its History, Theory and Development*. Chengdu, Sichuan: Sichuan University Press.

Zhang, J. (2007) 'A directional asymmetry in Chinese tone sandhi systems', *Journal of East Asian Linguistics*, 16(4), pp. 259-302.

Zhang, L.J. (2004) 'Awareness-raising in the TEFL phonology classroom: student voices and sociocultural and psychological considerations', *ITL International Journal of Applied Linguistics*, 145(1), pp. 219-268.

Remiszewski, M. (2005) 'Lingua Franca Core: Picture incomplete', in Dziubalska-Kołaczyk, K. and Przedlacka, J. (eds.) *English Pronunciation Models: A Changing Scene*. Bern: Peter Lang, pp. 293-308.

Zhou, W., Zhang, Y., Chen, H., Ji, X. and Shao, P. (2012) 'The acquisition of English tones by Mandarin EFL learners: A preliminary study', *Third International Symposium on Tonal Aspects of Languages*. Nanjing, China.

Zhu, X. (2003) *A Concise Guide to Contemporary English Pronunciation* Beijing: Foreign Language Teaching and Research Press

Appendix I: Demographic information of Chinese participants

ID	Group	Sex	Age	Residency in the UK (months)	English learning (years)	Birthplace (Province/ Municipality)	L1
1	Control	F	25	10	10	Henan	Northern Mandarin
2	Control	F	24	10	10	Zhejiang	Wu
3	Control	F	24	7	10	Zhejiang	Wu
4	Praat	F	24	8	10	Jiangsu	Mandarin
5	Audacity	F	28	7	10	Xinjiang	Mandarin
6	Praat	F	24	9	10	Guangdong	Mandarin
7	Praat	F	24	10	12	Guangdong	Mandarin
8	Audacity	F	25	7	10	Shandong	Mandarin
9	Audacity	F	31	8	14	Liaoning	Mandarin
10	Audacity	F	26	7	10	Henan	Mandarin
11	Praat	F	22	7	10	Beijing	Mandarin
12	Praat	F	24	8	10	Jilin	Mandarin
13	Audacity	F	24	9	10	Shaanxi	Northern Mandarin
14	Audacity	F	24	10	10	Zhejiang	Wu
15	Audacity	F	27	8	11	Beijing	Mandarin
16	Praat	F	23	10	10	Zhejiang	Mandarin
17	Audacity	F	23	8	10	Shanxi	Mandarin
28	Praat	F	25	2	10	Hebei	Mandarin
29	Praat	F	20	1	8	Shandong	Mandarin
30	Audacity	F	22	4	10	Anhui	Mandarin
31	Praat	F	22	1	10	Hubei	Southwestern Mandarin
32	Audacity	F	25	1	10	Hubei	Southwestern Mandarin

33	Praat	F	23	1	10	Sichuan	Southwestern Mandarin
34	Praat	F	24	15	10	Xinjiang	Mandarin
35	Praat	F	22	1	10	Anhui	Eastern Mandarin
36	Audacity	F	24	3	10	Guangxi	Mandarin
37	Audacity	M	23	1	10	Xinjiang	Mandarin
38	Audacity	F	23	1	10	Hunan	Xiang
39	Praat	F	25	1	12	Sichuan	Southwestern Mandarin
40	Audacity	F	24	1	10	Shanxi	Mandarin
41	Praat	F	23	1	10	Jilin	Mandarin
42	Audacity	F	23	1	10	Hunan	Xiang
43	Audacity	F	23	13	10	Hubei	Southwestern Mandarin
44	Praat	F	23	1	10	Shandong	Mandarin
45	Control	F	23	4	10	Zhejiang	Mandarin
46	Control	F	22	3	7	Shanxi	Northern Mandarin
47	Praat	F	22	2	7	Sichuan	Southwestern Mandarin
48	Praat	F	22	2	10	Hebei	Mandarin
49	Control	F	19	2	6	Shanghai	Mandarin
50	Control	F	23	3	10	Guangxi	Mandarin
51	Control	F	34	2	15	Henan	Northern Mandarin
52	Control	F	24	3	10	Jiangsu	Mandarin
53	Control	M	27	35	12	Shanxi	Mandarin
54	Control	M	24	15	10	Shanghai	Mandarin
55	Control	F	22	11	8	Anhui	Mandarin
56	Control	F	32	36	14	Liaoning	Mandarin
57	Audacity	F	21	13	9	Zhejiang	Wu

58	Praat	F	22	3	10	Henan	Northern Mandarin
59	Control	M	26	36	10	Neimenggu	Mandarin
60	Control	F	24	2	10	Guangdong	Cantonese
61	Control	F	23	8	10	Anhui	Mandarin
62	Audacity	F	22	3	8	Sichuan	Southwestern Mandarin
63	Praat	F	24	18	10	Taiwan	Mandarin
64	Audacity	F	22	2	10	Neimenggu	Mandarin
65	Praat	F	22	3	10	Hainan	Mandarin
66	Audacity	F	23	14	10	Hubei	Mandarin
67	Control	F	24	3	10	Jiangxi	Mandarin
68	Control	F	24	3	8	Heilongjiang	Mandarin
69	Control	M	24	2	10	Shandong	Mandarin
70	Control	F	22	1	6	Hubei	Mandarin

Appendix II: Pre-training Questionnaire

Thank you for participating in this questionnaire! It contains two parts. Part I aims to review your previous experience of English pronunciation training back in China and your own attitudes towards pronunciation. Part II is to survey your self-claimed knowledge of English prosody. The whole questionnaire will take approximately 10 minutes. Your answers are confidential; they will only be referred for academic purposes.

Part I:

1. Have you ever received English pronunciation training in China? (Please tick the format(s) of the training you have received. You can tick MORE than one.)
   a. Compulsory modules specifically on pronunciation
   b. Optional modules specifically on pronunciation
   c. No specific pronunciation training, but an integral part of the general English language modules I attended
   d. Extracurricular training in pronunciation (including private tutor, language training centre, e.g. New Oriental)
   e. Pronunciation as an integral part of extracurricular training in English
   f. Workshops on pronunciation
   g. I have never received any training in English pronunciation
   h. Other (please specify)

2. What percentage of the pronunciation training that you have received has focused on prosody (韵律, e.g. stress 重音, rhythm 节奏, intonation 语调…)?
   a. Less than 20%
   b. 20%~50%
   c. 50%~80%
   d. More than 80%

3. Can you recall what kind of English model(s) you have received in the training? (You can tick more than one)
   a. Standard British English (also named Received Pronunciation, Queen's English, Oxford English, BBC English)
   b. General American
   c. Other varieties (e.g. Australian, Indian, Singaporean, etc.)
   d. Chinese-accented English
   e. I don't know what kind of English models I have received
   f. Other models (please specify)

4. How was the pronunciation training conducted? (You can tick more than one)
   a. Listening and imitating the Chinese teachers
   b. Listening and imitating the Native English teachers
   c. Using native speakers' audio recordings (including authentic radio programs)
   d. Watching TV programs or films
   e. Assisted by computer programs (please specify what program(s) you've used)

5. Did you find the training helpful for improving your pronunciation?
   a. Yes
   b. No
   c. Not sure

6. Since you have ticked "yes", in what aspects have your pronunciation improved? (You can tick more than one)
   a. Consonants 辅音
   b. Vowels 元音

c. Lexical stress (词重音) (e.g. `present is a noun, pre`sent is a verb (` marks the stressed syllable))

d. Linking 连读 (e.g. “here are” read as [hiə.rə] instead of [hiə. ɑ])

e. Weak forms 弱读

f. Rhythm (or rhythmic beat) 节奏 (in longer utterances, the times from each stressed syllable to the next tend to be the same)

g. Intonation 语调

h. Other (please specify)

7. Do you think pronunciation is important for your English learning?

a. Yes, it is more important than other aspects, e.g. grammar, vocabulary.

b. Yes, it is equally important with other aspects, e.g. grammar, vocabulary.

c. Yes, but it is not as important as other aspects, e.g. grammar, vocabulary.

d. No, it is not important.

e. I am not sure.

Please explain why.

8. What aspects of your pronunciation do you think need to be improved? (You can tick more than one)

a. Consonants 辅音

b. Vowels 元音

c. Lexical stress (词重音) (e.g. `present is a noun, pre`sent is a verb (` marks the stressed syllable))

d. Linking 连读 (e.g. “here are” read as [hiə.rə] instead of [hiə. ɑ])

e. Weak forms 弱读

f. Rhythm (or rhythmic beat) 节奏 (in longer utterances, the times from each stressed syllable to the next tend to be the same)

g. Intonation 语调

h. Not sure

i. Nothing needs to be improved

j. Other (Please specify)

9. What kinds of English models would you prefer to learn? (You can tick more than one)

   a. Standard British English (also named Received Pronunciation, Queen's English, Oxford English, BBC English)

   b. General American

   c. Other varieties (e.g. Australian, Indian, Singaporean, etc.)

   d. I don't care as long as I am fully understood by my listeners

Can you explain why?

Part II: Please choose ONE option that most reflect your knowledge of each statement.

Statement	**Agree**	**Partly agree**	**Disagree**	**I don't know**
Intonation has three fundamental elements: phrasing (or tonality), accentuation (or tonicity), and tone.				
I know how to manipulate phrasing (or tonality) to make my speech fluent, coherent and unambiguous.				

I can hear how different phrasings (or tonalities) change the meaning of a sentence.				
I know how to manipulate accentuation (or tonicity) to emphasize or contrast the topic/focus of my speech.				
I can hear the focus/topic in other's speech by the placement of accentuations (or tonicity).				
I know how to change the tone of my speech to express my intentions.				
I can hear the intentions of a speaker's speech by his/her tones.				
I know how important intonation is in oral communication, particularly when talking with native speakers.				
I feel confident about my intonation.				

Appendix III : Stimuli for the comprehension experiment

Pre-test:

1. The guys in the car who were hungry ate some sandwiches.

Meaning:

Y: All of the guys in the car were hungry, and they ate sandwiches.

N: Not all of the guys in the car were hungry, and only the hungry ate sandwiches.

2. We're planning to fly to Italy.

In response to:

Y: Are you going to take a train to Italy?

N: Are they planning to fly to Italy?

3. Dawn: Unscrew the cylinder head.

Tom: Right.

Tom implies that:

Y: "I will."

N: "And what's next?"

4. She didn't do it because she was tired.

Meaning:

Y: She did it, but for some other reason.

N: She didn't do it. Here's why: she was tired.

5. Customer: Do you sell stamps?

Salesman: We do.

The salesman:

Y: confirms that they sell stamps.

N: may indicate that they've sold out the stamps.

6. Jack: Sophie's brought her friend along.

Sue: Who?

Sue is asking:

Y: Who has brought her friend along?

N: Which friend has Sophie brought along?

7. I won't tell anyone.

Meaning:

Y: I will tell no one.

N: I only tell a few people.

8. James: I was thinking of organising a collection for cancer research.

Sue: Well, I'll make a donation if you do.

Sue means that:

Y: If you organise it, I'll make a donation.

N: If you donate, I'll donate, too.

9. I do.

In response to:

Y: Do you smoke?

N: Who likes spinach?

10. You can have cheese salad or quiche.

Meaning:

Y: You have 3 options: cheese, salad, and quiche.

N: You only have 2 options: cheese salad, and quiche.

11. It’s snowing, isn’t it?

Meaning:

Y: I’m pretty sure it is.

N: I can’t see, I’m not sure.

12. It wasn’t under the table.

In response to:

Y: Where was it?

N: Was it under the table?

13. Look at the shoes she’s wearing.

Meaning:

Y: Look at the shoes she’s wearing, not the shirt.

N: Look at the shoes she’s wearing, not the shoes on the rack.

14. The water evaporated naturally.

Meaning:

Y: The water naturally evaporated, not through human intervention.

N: The water evaporated, as you would expect.

15. The king and the queen wearing ceremonial robes stepped out of the carriage.

Meaning:

Y: Both the king and the queen were wearing ceremonial robes.

N: Only the queen wearing ceremonial robes.

16. I’ve washed and ironed the clothes.

Meaning:

Y: I've got washed, and I've ironed the clothes.

N: I've done washing the clothes and ironing them.

17. He has a duty to perform.

Meaning:

Y: He must perform; that's his duty.

N: He must perform a duty.

18. Where are you from Bill asked Jim?

Meaning:

Y: 'Where are you from?' Bill asked Jim.

N: 'Where are you from, Bill?' asked Jim.

19. I'm singing, too.

Meaning:

Y: Not only are other people singing, so am I.

N: I am not only doing something else, but also singing.

20. On the Bank Holliday.

In response to:

Y: When are you going to the Angel of North?

N: Are you going to the Angel of North after the Bank Holliday?

Post-test

1. I just don't want anything.

Meaning:

Y: I want nothing.

N: I just want a few things.

2. Mike: There's something I must do.

   Sue: What?

Sue is:

Y: asking for a repetition.

N: asking what things Anna must do.

3. Daniel: Linda's thinking of organising an exhibition of ancient potteries.

   Jane: Well, I'll make a donation if she does.

Jane will make a donation if:

Y: Linda organises the exhibition.

N: Linda donates.

4. She can.

In response to:

Y: Can she help me put the books on that shelf?

N: Can you help me put the books on that shelf?

5. Mary is going to invite Peter, too.

Meaning:

Y: Someone else will invite Peter as well, not just Mary.

N: Mary will invite many people, not only Peter.

6. Where do you usually swim Jean asked Mary.

Meaning:

Y: 'Where do you usually swim?' Jean asked Mary.

N: 'Where do you usually swim, Jean?' asked Mary.

7. I have a book to read.

Meaning:

Y: I have to finish reading a book.
N: I don't need to write a book, just to read one.

8. Henry has written and performed a new monologue.

Henry has:

Y: written something, and he has performed a new monologue.

N: written a new monologue, and performed it.

9. The flags are red white and blue.

Meaning:

Y: Each flag has three colours: red, white and blue.
N: Some flags are red, some are white, and some are blue.

10. The youths who were wearing jeans weren't allowed into the restaurant.

Meaning:

Y: That group of youths weren't allowed into the restaurant.

N: All youths who were wearing jeans weren't allowed into the restaurant.

11. It's not right, is it?

Meaning:

Y: I'm sure it's not right.
N: I'm not sure, I'd like your views.

12. I bought it after Christmas.

In response to:

Y: When did you buy this?

N: Did you buy it before Christmas?

13. What about the book that you were writing?

Meaning:

Y: What about the book, not the script you were writing?

N: What about the book you were writing, not the book you were reading.

14. The cheese had gone mouldy naturally.

Meaning:

Y: The cheese had gone mouldy, as you would expect.

N: The cheese had gone mouldy because it was too old.

15. The old man and woman are playing golf.

Meaning:

Y: Both the man and woman are old.

N: Only the man is old.

16. Mike: Can we fix a date for the meeting?
    Celia: We could do it on Monday.

Celia is indicating that:

Y: “but maybe you won’t be available.”

N: “I know you’ll be free. Let’s meet up.”

17. No, I adore dogs.

In response to:

Y: Do you adore cats?

N: Do you object to dogs?

18. I'd like lamb and rice.

In response to:

Y: What about you?

N: What would you like with your rice?

19. I'm not going to come to the party because Sophie invited me.

Meaning:

Y: I'm going to the party, but for some other reasons.

N: I'm not going to the party, because Sophie invited me.

20. Jill: Do you know Peter?

Mary: Yes, I do.

Mary indicates that:

Y: I know him very well.

N: I just know this guy, but not like friends.

Appendix IV: Intonation instruction materials

英语语调

Session 1: Introduction

Yiling Chen (Jane)
IPhD in Phonetics and Phonology
School of Education, Communication and Language Science
Newcastle University
E-mail: y.chen21@Newcastle.ac.uk
Mobile: 07754608758

1

什么是语调?

声带(震动 ⟶ F0 (基频), 感知: pitch (音高)
仅仅只是音高运动吗?
Halliday的 3Ts 语调系统:
Tonality (调群切分)
Tonicity (调核分布)
Tone (调核音调)

2

WHAT'S YOUR INTONATION LIKE?

Monotone?没有音高起伏

Stress every syllable? 每个单词都重读

No proper pause, or pause too often? 停顿不当

Change tones on improper place? 语调变化位置不当

3

语调到底重不重要?

实证研究:

1. Chinese TA in American classroom failed to build up an interactive context (Pickering, 2004)

2. Prosody (韵律) rather than consonants and vowels accounts more for foreign accent (Munro & Derwing, 1995)

3. Improvement in intonation (prosody) contributes to greater intelligibility of learners' speech perceived by native speakers (Elliott, 1997)

4. Focus on most rewarding aspect of pronunciation (Munro, 2008)

4

ENGLISH INTONATION: AN INTRODUCTION, BY J. C. WELLS, 2006

"Native speakers of English know that learners have difficulty with vowels and consonants. When interacting with someone who is not a native speaker of English, they make allowances for segmental errors, but they do not make allowances for errors of intonation. This is probably because they do not realize that intonation can be erroneous."

"After all, almost any intonation pattern is possible in English, but different intonation patterns have different meanings. The difficulty is that the pattern the learner use may not have the meaning s/he intends. Speakers of English assume that – when it comes to intonation – you mean what you say. This may not be the same as what you think you are saying."

5

什么是韵律?

Prosodic (aka. **Supra-segmental**)

Pitch (音高)
Loudness (响度)
Tempo (speech rate) (语速节奏)
} make up the rhythm of speech

Intonation 语调, 往大了说其实包含上面所有!

6

HALLIDAY 的语调系统—3TS

Tonality （调群切分）

the division of the spoken material into chunks, also known as Intonation Phrase (IP)

e.g. Because I love languages I'm studying intonation.
|When I've finished this course, | I'll know a lot more about it.

7

HALLIDAY 的语调系统—3TS

Tonicity (nucleus)调核

the most important word (syllable) in an IP needs to be given special prominence

e.g. We're planning to fly to Italy.
We're planning to fly to Italy.
We're planning to fly to Italy.

8

HALLIDAY 的语调系统—3TS

Tone （调核调子）

pitch movements (stable or going down/up) hooked on nuclear syllables

e.g. It was remarkably `good.
It was remarkably `good.

9

语调短语的结构

The first accent before the nucleus if there's any is called the **onset** of an IP. From the onset to the syllable before the nucleus is the head. Syllables before the onset is the pre-head (no accented syllables). Any unstressed syllables after the nucleus is the tail.

(调冠) + (调头) + 调核 + (调尾)

We're | `planning to `fly to | `I taly.

NB. The boundaries of all elements do not necessarily coincide with word boundaries!!

10

语调短语

It was re`markably `good.

I'm `sure!

You `mustn't `worry.

11

总结：

"Speakers of English repeatedly face these 3 types of decision as they speak, namely 1. how to break the material up into chunks; 2. what is to be accented; and 3. what tones are to be used. These linguistic intonation systems are known respectively as tonality, tonicity, and tone. We refer to them as the three Ts."

--Wells (2006:6)

12

Paste

语调的功能

The **attitudinal function** (调子)

The **grammatical function** (调群 + 调子)

The **focusing** (also called accentual or informational) **function** (调核)

The **discourse (or cohesive) function** (调子)

The **psychological function** (调群)

The **indexical function** (所有) （母亲，律师。。）

13

调子：音高的起伏

陈述句一定以降调结束？ ✖

There's no simple predictable relationship between sentence type and tone choice. ✔

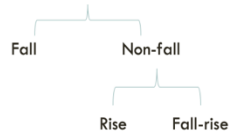
必须掌握的三点：

1. How many types of tones in English?
2. How to realise these tones?
3. What's the meaning of these tones?

14

英语调子的种类

两种基础调型：



15

调型赏析

-Whose is this?
-Mine.

-This money is John's.
-Whose?

-You said it was Black.
-White.

-Where do you come from?
-Spain.

-I'd like some chicken, please.
-Grilled?

-You promised it for Thursday.
-Friday.

Falls **Rises** **Fall-rises**

16

调子往哪儿放？

从调核开始：

降调：高开低走 (downward movement).

e.g. -`Where do you `come from? (tails in, fall extends)
-Spain. (no tails)

17

降调的实现

"There is very often a step up in pitch as we reach the beginning of the nuclear fall. Do not let this mislead you into thinking that the tone is rising(-falling)"!

e.g. How `very `strange!

`Then we'll see what happens!

18

英语语调

Session 2: Tone (Cont.)

1

RECALL 回顾

- **Three Ts** （三 **Ts** 系统）
  - Tonality (chunking/grouping): the smallest information **chunk**
  - Tonicity (tonic accent placement): the **focus** of a chunk
  - Tone: the pitch movement of accents
- **IP structure** (语调短语结构)
  - (pre-head) + (head) + **nucleus** + (tail)
- **Tone: fall** （语调——降调的形态）

2

RISES 升调

升调的语音实现?

- Starts with a relatively low pitch and ends with a relatively high pitch.
  起点：低——中　　终点：中——高
- If the nucleus is on the last or only syllable in the intonation phrase, then the rise takes place on that syllable. 如果在只有一个音节的调群里或调核在调群的最后一个音节上，整个升调就在这个音节内完成。
  e.g. Who?　　You want to talk to who?

3

升调的语音实现

- If there's a tail, the rising pitch movement will spread through the nuclear syllable and **ALL** the following syllables. This means that the last syllable is the highest pitched, even though it is unaccented. 如果有调尾的存在，那么升调会从调核开始一直延续到最后，也就意味着最后一个音节的音调最高。
- 升调的感知（母语者：低开处为最重读——中国人：高处最重读）
  e.g. Chicken?
  All of us?
  What did you say her name was?

NB. Usually there's a step down in pitch before the starting low point

4

降升调：难点

- **起点：**中——高　　**中点：**低　　**终点：**通常为中
- If the nucleus is on the last or only syllable in the IP, then the entire fall-rise movement takes place on that syllable. 同前，如果调群只有一个音节或调核在掉群最后一个音节，那么降升调会在这个音节内完成。
  e.g. Mine.
  I don't think it's mine.

5

降升调的语音实现

- If there's a tail, the pitch movement spread out over the nucleus and tail.
  如果有调尾的存在，那么降升调贯穿整个调核和调尾。
- The falling part takes place on the nuclear syllable, or between that syllable and the next. The rising part takes place towards the end of the tail and extends up to the last syllable of the IP.
  降调通常由调核和调核后一个音节完成，升调由剩下部分完成

e.g. Almost.
Mine, you mean.
Fortunately, | she wasn't.

6

调核声调的意思 (THE MEANINGS OF TONES)

- **Definitive fall:** what we say is potentially complete; we're confident, unreserved; signals finality. (特殊疑问句也是其中一种)

e.g. I'm delighted to meet you.
They're waiting outside.
She lives in Kennington as far as I can tell.
When did you arrive?

7

THE MEANINGS OF TONES

Implicational fall-rise: we typically state one thing but imply something further which is left unsaid—perhaps some kind of reservation or implication; signals a ***but***…to come but in a polite and indirect way. 有所保留和暗示

e.g. --Can we fix an appointment?
--I could see you on Wednesday.

NB. Implicitly or explicitly, it's up to you!

8

MORE MEANINGS OF FALL-RISE

- Say if we want to **correct** someone but do not want to be rude, use a falling-rising tone!

e.g. --She's coming on Wednesday.
--On Thursday.

Compare to this: --She's coming on Wednesday.
--No, on Thursday.

9

降升调的更多用法

- The fall-rise is often used in **negative statements** (positive part is left unsaid, but if said, use **definitive fall**):

e.g. I don't want to sound rude.
= I don't want to sound rude, but is that your dog?

e.g. --Why are you complaining?
--It's not just me.

=--It's not just me. Others are, too.

10

MORE OF NEGATIVE STATEMENTS

- **A special function: scope of negation (否定的范围)**

e.g. I won't eat \anything. Vs. I won't eat \/anything.
= I'll eat nothing. = I'll eat only certain things.

NB. It's ambiguous in written form!

11

DECLARATIVE QUESTIONS 陈述疑问句

- **Listen to the following dialogue:**

No one knows it's a question if you use falling tone!

--So there were three of them.
--Are you asking me or telling me?

More examples:
You'll be coming to dinner?
He took his passport?

12

YES-NO RISES 一般疑问句

- No matter if it is a positive or negative yes-no question.

e.g. Will your mother be coming?
Isn't she ready yet?

- **Pardon-question**

e.g. --I want to tell you something.
-- You what?/ Pardon?/ Sorry? (= I can't hear you.)

对比: What? (降调)

13

升调在特殊疑问句中的用法

- **Wh-questions:**

1. ***Encouraging rise*:**
   e.g. What's your name?
2. ***Interested rise*:**
   e.g. You can't go. Why **not**? Because I say so.
3. ***Echoic rise*:**
   e.g. –I'm moving to London.
   --**Where** are you moving to?

14

降调在一般疑问句中的用法

Insistent fall (repeating questions)

e.g. A: I'll ask you once more: Did you take the money?
B: No, I didn't.
A: Can you prove that?

A: Have you come far?
B: Sorry?
A: I said, have you come far.

15

Session 3: Tonicity (Part I)-where does the nucleus go
调核何去何从

英语语调

1

复习

1：什么时候用降调？名称？怎么实现降调？

Definitive fall, insistent fall

2：什么时候用升调？名称？怎么实现升调？

Yes-no rises, pardon-question,

3：什么时候用降升调？名称？怎么实现降升调？

Implicational fall-rise, scope of negation,

2

TONICITY（NUCLEAR ACCENT）调核

"The most important decision that the speaker makes in selecting an intonation pattern--to decide **where the nucleus goes**!"

！语调的核心部分：调核

❖Recall the structure of an IP?

调冠 + 调头 + 调核 + 调尾

The ***last accented*** lexical item

3

调核置放的规则（无标记位置UNMARKED）

1. On a lexically stressed syllable (won't change across the board)

词重音是潜在的语调重音！

So knowing the stress patterns of English is important! (查字典如果你不确定的话)

2. On the last lexical item (LLI) (incl. single content words and compounds)

Content words（实意词）: have meanings; can be defined; equivalents in other languages

e.g. 名词, 形容词, 大部分动词, 大部分副词

Function words（功能词）: whose meaning may need to be explained in a grammar rather than a dictionary; might no exact equivalent in other languages

e.g. 代词, 介词, 冠词, 助动词(do, have, has, shall, be…), 情态动词(助动词一种)

4

回顾：如何实现调核调子？

We put a nuclear tone on (or starting on) the nuclear syllable to make it the nucleus of an IP (Intonation Phrase).

！由调核所在音节开始

So how would you realise "tomorrow" as the nucleus?

e.g. tomorrow (falling)

tomorrow (rising)

tomorrow (falling-rising)

5

MORE EXAMPLES ON ONE-WORD NUCLEUS

--Have you been to Canada?
--Never.

--How did he look?
--Annoyed.

--What was the trip like?
--Indescribable.

--What won't he eat?
--Vegetables.

--What did she drink?
--Lemonade.

--How did she look?
--Ecstatic.

6

找出每句话的实意词：

e.g. She's done it! — done
Can you see her? — see
What are you looking at? — looking
Who was she talking to? — talking
I've just received a letter from her. — letter
Tell me about it. — tell
It's bigger than I imagined it would be. — imagined
Bill was talking at the same time as Jim was. — Jim

7

如果最后一个实意项是复合词呢?

Single-stressed
- **One-word:** e.g. `bedtime, `highlight, `keyboard, `grandmother
- **Two-word:** e.g. `credit card, `library book, `bus ticket, `running shoe, `high school, `water bill

Double-stressed
- **Name of people:** e.g. De,nise `Harris, ,Philip `Carr,
- **Name of roads and public places (except "…street")** e.g. Vic,toria `Road, ,Oxford `Avenue Eldon Square?
- **Name of institutions (hotels, schools):** e.g. ,Goldsmith's `College
- **Place or time in first element:** e.g. ,summer vo`cation
- **Material or ingredient in first element:** e.g. ,pork `chop beef stew?

Horrible !!!

8

有标记调核位置 (MARKED NUCLEAR ACCENT)

1. New information 新信息

We **accent** the new, and **de-accent** the old (given).

e.g. --Do you object to dogs? — Accent 'dogs', default
--No I adore dogs. — De-accent 'dogs', accent 'adore', new

--Shall we have the beef curry? — Accent 'curry', default
--No, let's have the prawn curry. — Accent 'prawn', new

9

有标记位置—新信息

Does it mean that we do not accent repeated words and accent all the new words? ✗

A. Synonym is a kind of old information.
e.g. –Shall we *wash the clothes*? --Oh I `hate doing the *laundry*.

B. Hypernym (上义词)is also a kind of old information.
e.g. *malaria* and `*other tropical diseases*.

C. Hyponym (下义词) is new information.
e.g. –Do you like *ball games*? --Well I'm quite fond of `*football*.

10

有标记位置—新信息

D. reusing the other speaker's words
e.g. --Would you like `coffee or `tea? --`Tea, please.
--I thought the cellos were `marvellous. --The `cellos?. I didn't.

E. Reusing your own words
e.g. When I say `stop, `stop!
I can't ac`cept it, I just can't ac`cept it.

11

OCCASIONS APART FROM THE LLI RULE

2. Contrastive items

e.g. A: Mary gave John a camera.
B: No, she didn't give it to him; he gave it to her.

Any word can be accented for contrast, e.g. pronouns, prepositions…

e.g. I'm writing a letter. What are you doing?
It wasn't under the table, but actually on it.
I can send a fax to him, but I can't receive one from him.

12

REMEMBER!

非实意项一般不作为调核除非它们被强调对比：

e.g. A: Have you hurt yourself? — Default accent
B: Yes I've cut myself. — Default accent/hyponym

He did it all by himself. — Contrastive—without help from others

13

调核置放—非实意项

The same applies to *someone, something, somewhere, somebody,* and their negative equivalents *anyone, anything, anywhere, anybody.*

e.g. Can I get you anything? — Default accent
His name was Jimmy, or Billy, or something. — Default accent
I won't tell anyone. — Default accent
I won't tell anyone. — Marked negative
I won't tell anyone. — I'll only tell a few people

14

HOW TO CONTRAST POLARITY (否-肯定极) AND TENSE

Sometimes you want to emphasize the polarity (the quality of being either positive or negative) of a verb, or its tense. The nucleus in both cases goes on an auxiliary or modal verb.

有标记的否定极: on the negative word (not) or the contracted (won't, etc.)

e.g. He thought I'd finished the essay, but actually I hadn't finished it.

--Have some more milk. --I don't want any more milk. ✗ — Accent 'want'—new lexical item after negation

15

HOW TO CONTRAST POLARITY AND TENSE

有标记的肯定极: on the modal or auxiliary verb that indicates positive polarity; or restatement of an existing polarity.

e.g. --You're not involved. --Oh I `am involved!
He promised he would finish it, and he actually did finish it.
--I don't like Beethoven. --Well I do. — When both subject and polarity are in contrast, accent the subject.

Compare: --I was thinking of organising a collection for cancer research.
--Well, I'll make a donation if you do. — =if you organise it
or --Well, I'll make a donation if you do. — =if you make a donation

16

英语语调

SESSION 4: TONICITY PART II

1

回顾

无标记调核的位置?

- 1. LLI (last lexical item) 最后一个实意项
- 2. on the lexically stressed syllable of LLI 最后一个实意项的首重音

有标记调核?

- 1. New information 新信息
- 2. Contrastive information 对比信息

2

A NEW CONCEPT: *FOCUS* 焦点

What is focus?

– the concentration of attention on a particular part of the message

Types of focus

1. Broad focus: 宽焦点

everything in your utterance is news/focus (so LLI applies)

2. Narrow focus: 窄焦点

something in your utterance is new/focus (non-LLI applies, either the informative, or the contrastive, or emphatic)

3

EXAMPLES OF BROAD/NARROW FOCUS

e.g. --What happened next? — Broad focus, LLI rule applies, nucleus is on "***next***"

--Everyone burst out laughing. — Also broad focus, LLI rule applies, nucleus is on "***laugh***"

--Who brought the wine? — Broad focus, LLI rule applies, nucleus is on "***wine***"

--I think it was Mary that brought the wine.

Narrow focus, non-LLI rule applies, nucleus is on "***Mary***"

4

BROAD VS. NARROW

I've just been talking to a friend of mine. — Broad focus, LLI rule applies, nucleus is on "***friend***". If you accent "mine", you must contrast other's friends!

I think we all ought to help one another. — Broad focus, LLI rule applies, nucleus is on "***help***". Reciprocal pronouns are rarely contrastive or new

Have you hurt yourself? — Broad focus, LLI rule applies, nucleus is on "***hurt***". If on the reflexive pronoun "yourself", there must be a common sense that somebody else is hurt as well.

Can I get you anything? — Broad focus, LLI rule applies, nucleus is on "***get***". Indefinite pronoun is usually not contrastive, only when you definitely bring them to.

5

BROAD VS. NARROW

❖ ***Where is the nucleus?***

--Is Peter coming?	--Yes, he is.	*Yes; is*, narrow focus/emphasize
	--Yes, Peter is coming.	
--Who's coming?	--Peter is.	*Peter*, narrow focus/emphasize
	--Peter's coming	
--Where's Peter?	--He's coming.	*coming*, new/default
	--Peter's coming.	

6

IF ALL ARE FUNCTION WORDS? 一句话全是功能词咋办?

e.g. --Look at this button.
--What's it for? → **What's that button for?**

--That's Mary.
--Who's she with? → **Who's Mary with?**

--You know my essay? ***--Yes, what about it?***
--What do you want to say about it?

7

其他可作调核的功能词

- ***Too*** e.g. I'm singing, too. I'm singing, too.
- ***As well*** e.g. Could you give me some please, as well?
- ***Anyway*** e.g. This idea may not work, but let's try it anyway.
- ***Anyhow*** e.g. She doesn't smoke—not nowadays, anyhow.

8

EMPTY CONTENT WORDS

- ***Things, People*** e.g. What are you going to tell people?
  I keep seeing things.
- ***One, So*** (pro-form: to refer back to an idea, situation etc.)
  e.g. The train was crowded, so we caught a later one.
  If you're feeling unwell, just say so.

9

ADVERBS OF TIME AND PLACE OFF NUCLEUS表时间地点的副词

❖ Although adverbs in general are usually accented, adverbs and adverbial phrases of time and places are often not accented, otherwise they must be perceived as contrasting to shared knowledge.

e.g. I had an unexpected letter yesterday.

Did you see Big Brother on television last night?

--Everything Ok? --It's a bit hot in here.

10

FINAL VERBS AND ADJECTIVES句末动词和形容词

❖ If there's a verb at the end of a sentence or clause, it's usually de-accented, and the nucleus goes on a preceding noun.

e.g. How's the homework going?
I've still got an essay to write.
You ought to keep his mouth shut.
You need to keep the brush wet.

Compare: ***I've got some work to do.*** *Accent "work", the noun proceeds "to do"*
Tell me what to do. *Accent "do", no noun proceeds it*

11

EVENTS （一种特殊句型）

Sentences describing an event, where the verb is intransitive (不及物动词), is called event sentences. The nucleus is usually going on the subject, rather than the verb even though it carries new information.

e.g. The phone's ringing. — **The phone's exploded!**
The kettle's boiling.
Your house is on fire.
The sun's come out.

If the verb is predictable form the context, it doesn't need to be in focus

12

PHRASAL VERBS 动词短语

Format: verb + particle (动词+小品词)

- ***Intransitive:*** *作不及物动词用 （双重音，首重音在助动词上）*

e.g. Come in, sit down.
The next month she passed away.
The prisoner broke down.

Accent "in", "down", "-way", "down", nucleus on particles

例外：It was really pouring down. → Accent "pour-"

13

PHRASAL VERBS 动词短语

Format: verb + particle (动词+助动词)

- ***Transitive:*** *作及物动词用 （单重音，在动词上）*

no object (at the end): *What are you looking at?* → Accent the verb

Separated particle
- Noun: *Take your shoes off.* → Accent the noun
- Pronoun: *Take them off.* → Accent the particle

例外：look after, do without, bump into

14

HOW TO REALISE TONICITY? 调核的语音实现
注意：已经不光是调核调子的语音实现了

根据定义：

Tonicity = the most prominent part of an intonation phrase (information focus)

（突显）

Prominence is realised by
- Lengthening the duration 时长加长
- Increasing the amplitude/intensity 强度加强
- Articulating the vowels more precisely 元音饱满清晰
- Changing pitch most dramatically (pitch minima/maxima) 音高明显变化

Non-nuclear but stressed syllables might also have these features!

15

THE REALISATION OF NUCLEUS 调核的语音实现

1. Lengthening the duration 时长加长

Relative longer duration (注意：不是绝对时长) compared to other non-stressed syllables

Lengthen the nucleus, **or shorten** the non-stressed syllables.

e.g. The porter and I looked at each other and smiled.

Chinese learner's vs. Native speaker's

All syllables are in similar duration.

De-accent *and, at, each other, and; lengthen porter, I, looked, smiled.*

16

THE REALISATION OF NUCLEUS 调核的语音实现

2. Increasing the amplitude/intensity 强度加强

No problem for Chinese students!

❖ But decreasing the amplitude/intensity of unstressed words, the other way of making the nucleus prominent, is indeed a problem.

e.g. My dentist had just pulled out one of my teeth and had told me to rest for a while.

Chinese student's vs. Native speaker

De-accent *had, and, to* by minimising their intensity.

17

THE REALISATION OF NUCLEUS 调核的语音实现

3. Precise articulation of the vowels 清晰饱满的元音

= making the unstressed syllable less precise!

e.g. I had an unexpected letter yesterday.

Linking, /dən/

e.g. My dentist had just pulled out one of my teeth and had told me to rest for a while.

Elision, zero realisation of /h/

18

THE REALISATION OF NUCLEUS 调核的语音实现

Cont.

e.g. --Do you object to dogs? --No I adore dogs.

Weaken the vowel /ɒ/ into Schwa /ə/

19

THE REALISATION OF NUCLEUS 调核的语音实现

4. Changing pitch most dramatically (pitch minima/maxima)
音高明显变化（出现最高音或最低音）

e.g. On the way home, he asked Jenny if she had enjoyed the speech. To his surprise, she said she hadn't.

Chinese student's vs. Native speaker's

Accent *she* with a dramatic fall, making it focused (nucleus)

20

总结：

❖ These four aspects are involved in nucleus realisation, either all are or just some of them are, but pitch change is the priority.

感受下面这句话，当一个单词作为调核的时候和它没作为调核的时候，有什么样的差别？

e.g. We're planning to fly to Italy.
We're planning to fly to Italy.
We're planning to fly to Italy.

21

BONUS: CHECK THIS OUT!

Amy: Okay, let's get this Victorian party started. I've done some research about some traditional parlor games we can play.

Penny: Like what?

Amy: Well, there's a fun one called Ball of Wool. You take a ball of wool, and you put it right in the center of the table. Then people sit on opposite sides of the table and try to blow the wool off.

22

英语语调

Session 5: Tonality 调群切分

1

复习

- **调核置放的规则：**
  1. 无标记位置的规则是？
  2. 有标记位置的规则是？
     宽焦点 vs. 窄焦点
- **调核的语音实现方法：**
  1. 相对时长；2. 相对强度；3. 元音饱满度
  4. 音高明显变化

2

调群切分的三个关键问题

1. where do the boundaries between two successive IPs go? 调群边界应该在哪儿?
2. What determines the number of words that go into a single IP? 一个调群到底有多少个单词?
3. How to signal an IP boundary? 调群边界的语音指征?

3

大体上来说，

Intonation break corresponds to syntactic boundary, 即调群的切分基本符合句法结构

e.g. between successive sentences, successive clauses, sometimes between successive phrases, and occasionally between successive words, or even between successive syllables when special emphasis is needed.

e.g. Milk comes from cows, | and wool comes from sheep.

Milk | comes from cows.

Delicious, | cool | milk.

Bor|ing!

4

IP CAN DISAMBIGUATE AMBIGUOUS UTTERANCES
调群切分有消除歧义的作用

e.g. Help keep the dog off!

vs. Help! Keep the dog off!

You can have cheese salad or quiche.

Version 1 Version 2 Version 3

The competitors who finished first received a goody bag.

Version 1 Version 2

5

THE SIZE OF AN IP (INTONATION PHRASE)
调群的大小

- IP is not only the basic chunk for intonation purposes but also the *basic chunk for mental planning*.
- In our minds as we speak we plan *one IP at a time* and *a single focus domain*.
- So don't worry too much about this, as long as you bring your focus of your speech into accent, and speak out *one focus domain at a time*.

6

HOW TO REALISE IP BOUNDARY
调群的语音实现

***1. Pause*停顿 :** the most typical boundary marker of IPs

filled pause (填声停顿) **vs.** unfilled pause (无声停顿)

err, um, … | silence

There seems a relationship between the length of pause and boundaries: the more major the boundary, the longer the pause.

e.g. version 1 vs. version 2

Can you hear IP boundaries and the nucleus?

7

2. ANACRUSIS 起首轻音节

➢ Refers to a string of unstressed syllables at the beginning of an utterance (before the onset) with greater speed, and more reduced form. *调冠的出现是另外一个调群开始的表征*

e.g. He then asked me | how my brother was | and whether I `liked my new job in London.

8

3. SYLLABLE LENGTHENING
音节延长

➢The final syllable of an IP seems to be lengthened to indicate that another IP is going to start, and it's usually followed by a pause.

a. A natural relaxation before pause, or a substitution of a pause
b. Provides time to check the preceding group has been articulated correctly
c. Be given time to plan the following group

e.g. …where I go, | the phone goes. | No phone, |no Neil. |

Unstressed, but lengthened | Stressed, lengthened

9

4. PITCH RESET 非重读音节音高重设

➢A pitch change (in pitch level or in pitch direction) usually happens to stressed syllables, but it can also occur on unstressed syllables, and by so a new IP starts.

Here's a piece of authentic recording, in which Professor Robin Ferner is talking about the bad effects of laughter:

We found people who had had heart rhythm disturbance | which had stopped their heart, | we found people who had fainted, | we found people whose gullets had burst, | we found people who'd dislocated their jaws | or burst their lungs.

Unstressed, but lengthened | Compared to the previous pitch level

10

总结:

Tonality, just as tonicity and tone, is very personal. You can phrase your speech wherever you want. But if you don't mean to cut your speech up, then don't show up any features of the abovementioned, otherwise people might misunderstand you!

e.g. He was looking up the street.

Version one: =gazing up the street, "up" is a prep.

Version two: =searching for the street, "look up" is a phrasal verb

11

SYNTACTIC UNITS WHICH NORMALLY FORM A SEPARATE IP 符合句法结构的调群切分

***1. Co-ordinate clause* 并列成分**

e.g. Mary moved to Paris, | but John stayed in London. (sentence co-ordination)

John went to the pub | and ordered a beer. (verb phrase)

She's very tall | and very pretty. (adjective phrase)

His wife dresses very well | and very quickly. (adverbial phrase)

It's either in the fridge | or on the table.

Listen, **I've washed and ironed the clothes.**

Version 1 Version 2

12

CO-ORDINATE CLAUSE 并列从句

❖过短的并列成分通常不够成单独的调群.

e.g. He stopped and stared. (verb phrase)

He bought milk and cheese. (noun phrase)

fish'n' chips

strawberries and cream } Lexicalized elements

Bill and Mary (names of couples)

The Fox and Hounds (pub name)

13

SYNTACTIC UNITS WHICH NORMALLY FORM A SEPARATE IP符合句法结构的调群切分

***2. Listing items* 列举**

e.g. The flags are red, white and blue. (the flags are three colours)

The flags are red, | white, | and blue. (some flags are red, some white, and some blue)

❖*Letters or numbers, either one IP or multi-IPs based on the shared knowledge between the speaker and the hearer.*

e.g. Your reference | is A L W | 2 0 0 4 | 3 9 6 J.

14

SYNTACTIC UNITS WHICH NORMALLY FORM A SEPARATE IP符合句法结构的调群切分

***3. Subordinate clause* 从属句**

Please do not bring these words into focus or separate IP

e.g. I'll buy the fish'n' chips | when I go to the shops.

I told the new recruit to the company | that he was fired.

❖However, if the material before the clause is short, then the clause doesn't have to be isolated.

e.g. I think she's been sacked.

Compared to I think | she's been sacked. (sacked=fired)

15

SYNTACTIC UNITS WHICH NORMALLY FORM A SEPARATE IP符合句法结构的调群切分

***4. Sentence adverbials (speaker's attitude)* 句子副词**

They normally have their own IP, because the clause would be grammatically complete without them.

e.g. John went to the interview, | hopefully.

or Hopefully, | John went to the interview.

Others: technically, honestly, unfortunately, obviously, basically…

Compared to: John went to the interview hopefully.

16

SYNTACTIC UNITS WHICH NORMALLY FORM A SEPARATE IP符合句法结构的调群切分

***5. Cleft and pseudo-cleft sentences* 分裂句及假拟分裂句**

One constituent of the sentence is fronted and introduced by ***it is*** (or ***it was***, etc.) or by ***what***. This constituent is thus brought into focus and must carry the intonation nucleus. If there's further focused materials to follow, then there must be an intonation break.

e.g. **Cleft:** It's John | that I love. = *I love John. (or **ONE** IP)*

Pseudo-cleft: What he needs | is a bath. = *He needs a bath.*

Cut the construction into 2 halves

17

SYNTACTIC UNITS WHICH NORMALLY FORM A SEPARATE IP符合句法结构的调群切分

6. "is…is that" 句型

A widespread informal/formal construction in spoken English. The norm is having an IP after the first "is" which usually has a fall-rise nuclear tone.

e.g. The `thing is | is that she's pregnant.

My con`cern is | is that it's got too big | too quickly.

The `fact of the matter is | is that the `way it is run | is too complex.

18

SYNTACTIC UNITS WHICH NORMALLY FORM A SEPARATE IP符合句法结构的调群切分

***7. Parenthesis (lower pitch range)* 插入语**

- **Non-restrictive relative clause as a parenthesis**

e.g. –Who's Nikki? --She's my sister, | who lives in Canada.

Compare to this, --She's my `sister who lives in Canada.

- **Noun phrases in apposition 同位语**

e.g. `Barak Obama, | a `Democrat politician, | is intelligent.

- **Others (clause, sentence adverbials…)**

e.g. Mary, | you're not `going to believe this, | but `Jane is pregnant.

19

SYNTACTIC UNITS WHICH NORMALLY ***NOT*** FORM A SEPARATE IP 不符合句法结构的调群切分

Reporting clauses (wide used in daily speech, although abound in novels)

句末报告短语

e.g. | "What has happened?" she asked. |

| "Good heavens!" exclaimed his wife. |

| "And how can you explain that?" she asked. |

Compare: He's retired, I think. Vs. He's retired, | I think.

20

Slides

英语语调

Session 6: Put them altogether

1

TAG QUESTIONS 附加疑问句—反义

- Short yes-no questions tagged onto the end of a statement, command, or exclamation. In some cases, both tone and tonality matter.

e.g. The answer is twenty, | isn't it?
We could start with the kitchen, | couldn't we?
Yes-no rise for asking questions

The view is magnificent, | isn't it?
We've been here before, | haven't we?
What a surprise, | wasn't it?
Insistent fall for appealing for agreement

Reverse-polarity tags usually have their own IP

2

TAG QUESTIONS 附加疑问句—非反义

- **Constant-polarity** tags (copy tags) usually do not have their own IP, thus forming the tail of the IP. (Normally positive-positive) 通常指说话双方共享的信息，陈述部分用调核降调，附加部分以中/高调位结束

e.g. So you've qualified as a lawyer, have you?

Come over here a minute, will you?
Open the window, would you please?
命令是一种非反义疑问句.

3

SEQUENCES OF TONES 音调的排列顺序

❖***Leading and trailing tones***

- Many phrases or clauses do not stand alone, because they're not complete. They are attached to some other elements which are independent.

Before → Leading element → Leading dependent fall-rise

After → Trailing element → Trailing dependent rise

4

SEQUENCES OF TONES 音调的排列顺序

The independent part → Independent fall

e.g. After lunch | we could call on Mary. → Leading + independent

We're going to Spain | in August. → Independent + trailing

If I were you, | I'd rejected. → Fall-rise + fall

I'd rejected it, | if I were you. → Fall + rise

❖Fall-rise is a very characteristic tone of mainstream English (RP, GA)

5

TONE CONCORD 音调和谐

- If two or more successive intonation phrases are grammatically parallel, they can be signalled by using the same tone. 既排比结构可用同一调核调子来说

e.g. We found people who had had heart rhythm disturbance which had stopped their heart, | we found people who had fainted, | we found people whose gullets had burst, | we found people who'd dislocated their jaws | or burst their lungs.

- When *too, as well, either* form a separate IP, they usually exhibit tone concord.

e.g. If Martha wants to come, | as well, | we'll need a bigger car.

6

VARIETIES OF FALL 降调的种类

Recall the meaning of fall?

definitive, complete (independent), insistent (yes-no Qs), and reinforce (adverbial)

3 varieties of fall:

Have all these, but with addition

high fall (高降) High to low | **low fall (低降)** Mid to low | **rise-fall (升降)** Mid to high to low

7

MEANINGS OF FALLS 各种降调的意思

High fall: I'll be staying for a month. (excited, enthusiastic)
I don't know what to do. (hopeless)
Come and have dinner with us. (warm invitation)

Low fall: I'll be staying for a month. (factual, objective)
I don't know what to do. (factual)
Come and have dinner with us. (serious order)

More emotion involved → Less emotion involved

8

MEANINGS OF FALLS 各种降调的意思

Rise-fall

- *The speaker is impressed (reinforced by breathy voice, gossipy)*
  - e.g. –She came top of the class. --Did she just! | Well well.
  - --Jane has left her husband. --Terrible, | isn't it?
- *The speaker shows challenge or even disapproval.*
  - e.g. –I need a break. --Don't we all!

9

简单调头(只含一个重音)

High level: (高平) Well `make up your mind. (remain high level pitch) + rise/fall/fall-rise

High falling: (高降) Well `make up your mind. (mid-high to mid) + fall-rise

Low level: (低平) Well `make up your mind. (remain low level pitch) + rise

Low rising: (低升) Well `make up your mind. (low to mid) + fall

10

调头调核组合的意思

降调调核的调头通常为高调

Head	+ nucleus	
high	high-fall	Emotional involvement
high	low-fall	Unemotional and factual
low-rise	high-fall	Protest (抗议)

e.g. We `handed them in yesterday. (high + high-fall)
We `handed them in yesterday. (high + low-fall)
We `handed them in yesterday. (low-rise + high-fall)

11

调头调核组合的意思

调头 + 调核

Low-level → rise → Defensive, grudging, or generally non-supportive

e.g. –The top's fallen off. --I `don't suppose it matters.
--Her performance was dire! --Oh it `wasn't as bad as that.

12

调冠

Low pre-head: neutral, unmarked 低调冠，无标记
e.g. Well `make up your mind.

High pre-head: marked, emphatic (usually very short, no more than two/three syllables) 高调冠，有标记
e.g. Well `make up your mind.
(usually higher than the rest of the IP)

13

KNOW HOW TO DO ANACRUSIS? 起首轻音节要怎么实现呢?

- Unaccented words or phrases at the beginning of the utterance, served as pre-head.

e.g. Yes I'll `do it tomorrow.
(high + high level head + high fall)
I mean I'm `not trying to criticize, | but…
(high + high fall head + fall-rise)
And | the `sad thing was | that…

14

NO PRE-HEAD CONSTRUCTION 零调冠

- **Onset on a function word** 调头首音节由功能词（语法词）充当

1. *疑问词:* `Who wrote the report?
`This is the officer | who `wrote the report. (Relative wh- words are usually de-accented)
2. *指示词:* `This boyfriend of yours, | `what does he do for a living?
3. ***"there"*** *作为地点副词开头:* `There he sat, drinking his beer.
(Pronoun, not accented) There's `nothing we can do.

15

总结:

- ❑ Speak as you think; think as you speak.
- ❑ Accent important (new, contrastive, emphatic) information; de-accent the unimportant.
- ❑ The unimportant part can be heavily reduced with fast speed.
- ❑ Do not pause if your minds are flowing, but make sure your information chunk is clear.
- ❑ Do not always put a fall on accented words, fall-rise and rise are more likely to attract other's attention.

16

Appendix V : Practice materials for the Praat group

Session 1

I. Introduction

*What is **Praat**?*

A Dutch word for “talk” or “speech”, a flexible software for speech analysis and synthesis designed by Paul Boersma and David Weenink at the Phonetic Sciences department at the University of Amsterdam.

目前语音科学研究应用最广泛的声学分析软件，其功能包括录音，声音数据标注，声学指征数据分析，语音合成等。其界面简单，直接，易操作。本试验只涉及录音，语调数据观察，分析和对比等满足语调学习的基本操作。对其他方面有兴趣的同学请在本试验结束后索取其他信息。

II. Start-up

重要概念 (Terminology)

１、采样 (Sampling)

由于声音为模拟连续信号，而计算机只能处理数字离散信号，因此要使用计算机来分析和处理声音，就需要经历模数转换过程[Analog to Digital Converter，即 ADC]，即将模拟的连续信号转换为数字离散信号。采样就是按照一定的时间间隔从模拟连续信号提取出一定数量的样本来，其样本值用二进制码 0 和 1 来表示，这些 0 和 1 便构成了数字音频文件，其过程实际上是将模拟音频信号转换成数字离散信号。

２、采样率 (Sampling rate)

采样率表示了每秒对原始信号采样的次数。显然，在一秒中内采样的点越多，获取的信息越丰富，为了复原波形，一次振动中，至少得有 2 个点的采样，要想使采集到的信号不失真，采样频率规定至少为语音频率的 2 倍，因此要得到一个频率为 10000 赫兹(Hz)的声音，则其采样率至少得大于 20000 赫兹。采样频率越高，数字信号的保真度越高，但同时占用的存储空间也越大。如果采样率低于高频成分频率的两倍，则会产生低频失真、信号混淆现象。

３、采样精度 (Sampling size)

采样精度就是指存放一个采样值所使用的比特数目。当用 8 个比特 (bit)（采样精度为 8 位）存放一个采样值时，对声音振幅的分辨等级理论上为 256 个，即 0 至 255；当用 16 个比特（采样精度为 16 位）存放一个采样值时，对声音振幅的分辨等级理论上为 65536 个，即 0 至 65535。如果您将采样精度设置为 16 位，计算机纪录的采样值范围则为-32768 至 32767 之间的整数。采样率和采样精度的值越大，记录的波形更接近原始信号，但同时占用的存储空间也越大。

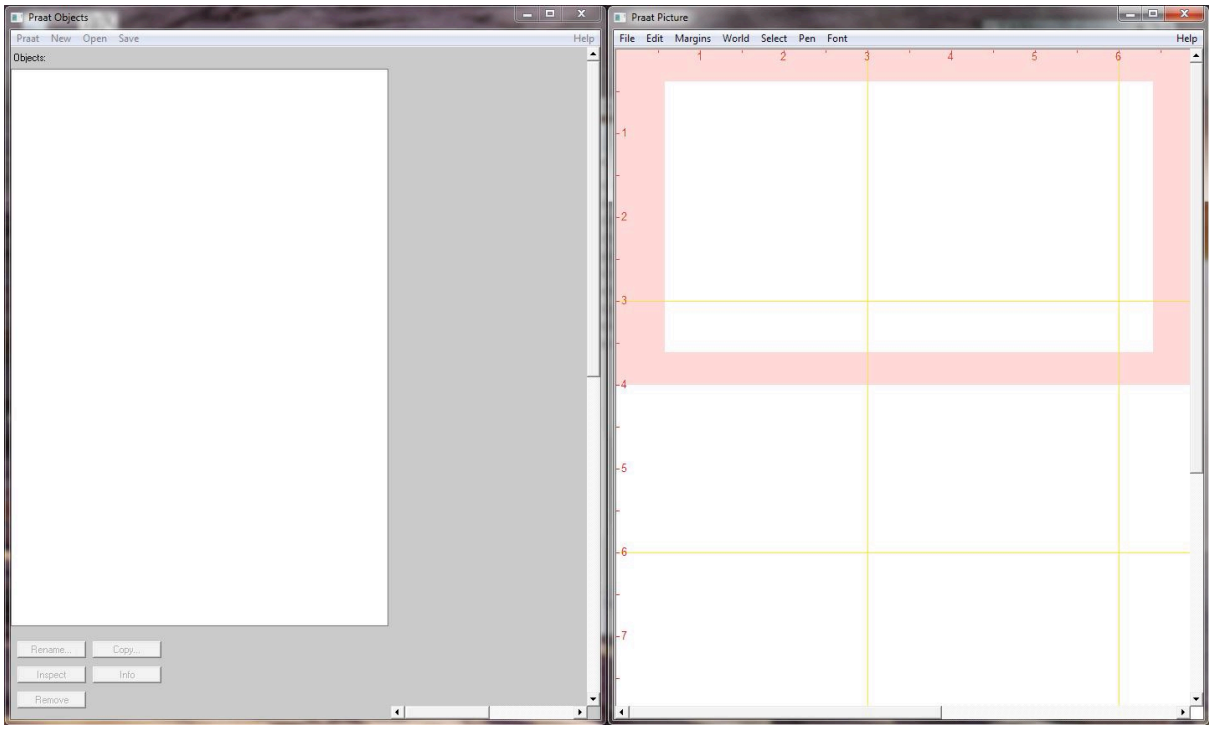
４、声道 (Sound channel)

声道指输入或输出信号的通道。通常用多声道来输入或输出不同的信号。如果只需录制一个位置的一种信号时，只要使用单声道就可以了。

声音采集 (Recording a sound file)，保存，及打开

1. 登陆你自己的学校账户。离开教室前，记得登出 (log off)。

2. 双击桌面上的 Praat 图标，然后就会出现下面两个窗口: (图 1)



左边是目标窗口 (Praat Objects)，所有打开的文件 (声音及标注文件) 以列表的方式出现在这里，以便对其进行观察，分析和保存。右边是绘图窗口 (Praat Picture)，本试验用不着，请点击右上 x, 关闭该窗口。

3. 戴上耳麦，将话筒置于你嘴巴正前方偏左一点点大概 5 厘米的位置。

4. 点击目标窗口上的 new > record mono sound, 会弹出一个录音窗口 (SoundRecorder)。默认采样率是 44100Hz, 足够采样到人类声音，确保这个采样率被选取 (见图 2)。

5. 准备就绪后，点击 Record, 对着话筒念出: “测试，测试”，注意观察中间的信号强度指示带 (Meter)，保持绿色说明强度合适，往上到达 2/3 处出现黄色时，说明信号过强开始扭曲，红色则完全扭曲。所以念的时候尽量保持不要出现黄条或刚刚出现黄条。念完后点击 Stop，然后点击 Play 检查声音是否正常录入，音质是否好。重复录音直到你自己满意为止。每次新的录音就会自动删除上一次的录音，除非你对上次录音有所保存 (见第 6，7 点) 。注意默认每次录音时间不能超过 90 秒，但最好保持在 3 秒内！

6. 当你对某次录音满意时，在 “Untitled” 处重新命名这个声音文件, 比如“testing”, 然后点击 Save to list and close。

7. 你重新命名的这个声音文件此时就会出现在目标窗口中，见图 3 左。将这个声音文件保存在你自己的 H 盘中以便以后提取或分析：点击 Write > write to WAV file… (声音文件的格式为.wav)。要提取文件，则点击 Read > Read from file…，双击你的文件，然后该文件就会出现在任务窗口中, 见图 3 右。

图 2：

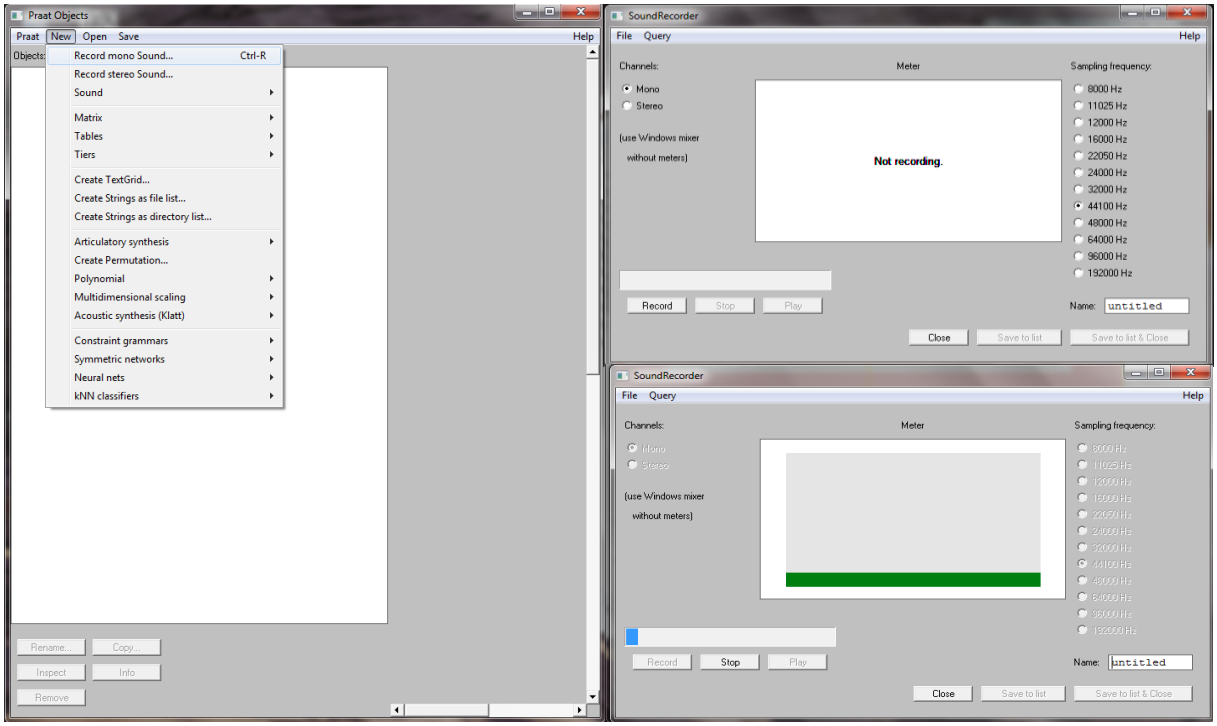
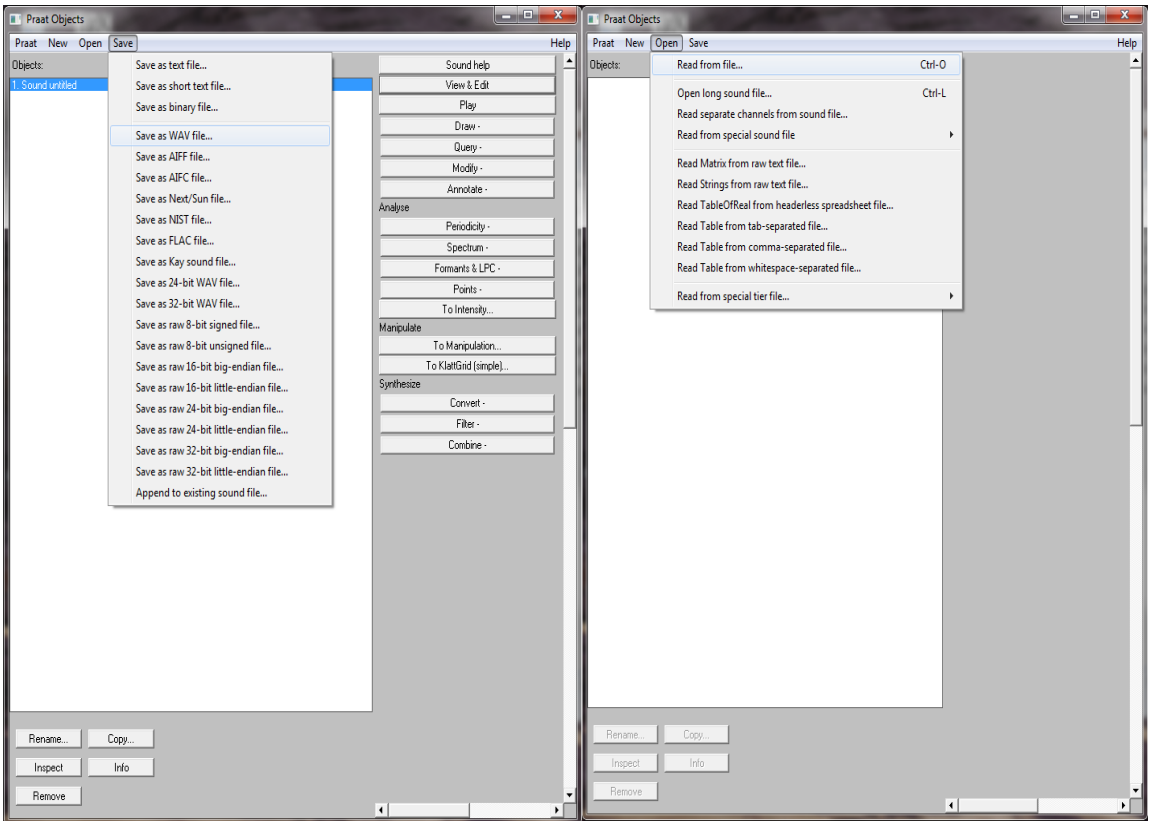


图 3：

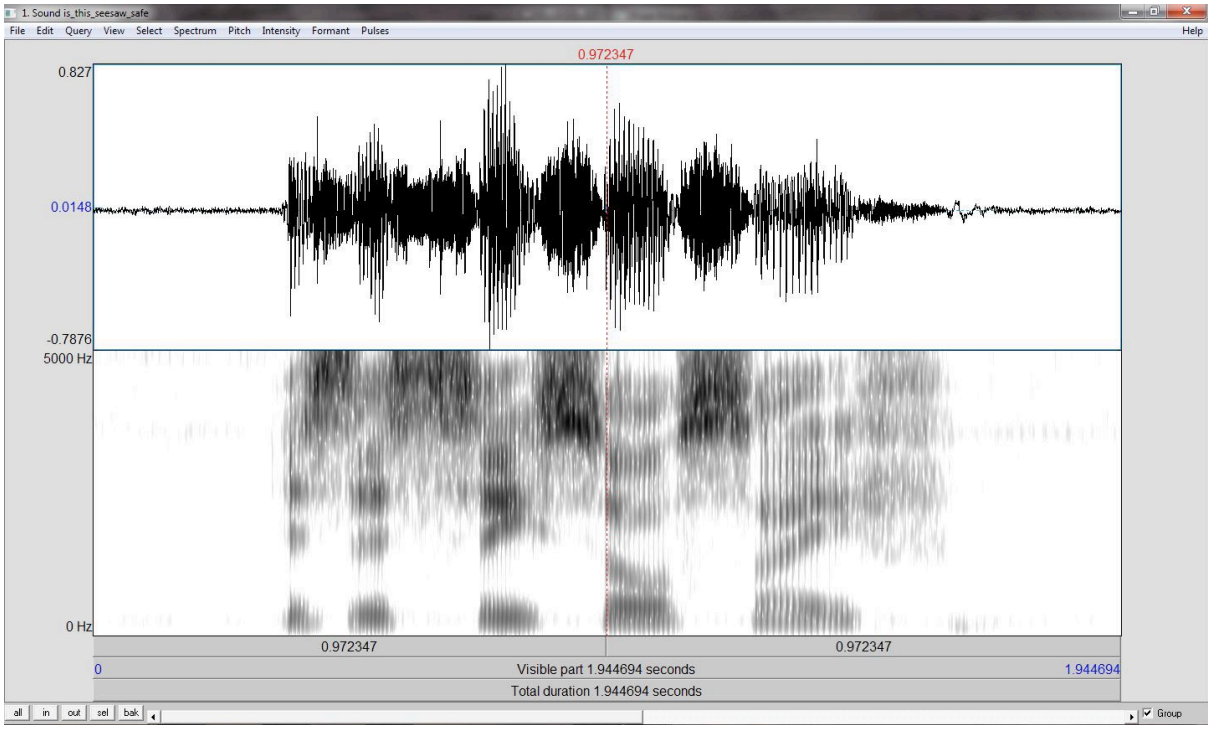


查看&编辑声音文件

(View & Edit sound files)

1. 在目标窗口中选取你保存的声音文件, 如 “testing”, 然后点击右边菜单中的 View & Edit, 一个新的编辑窗口就会弹出，此图就展示了该声音文件的所有声学信息 (见图 4)。可点击右上角放大该窗口至全屏，以便更好的观察。

图 4：



该窗口被分成了上下两个部分，上部分叫声波图 (waveform)，下部分叫频谱图 (spectrogram)。两部分其实都是三维的，横轴表示时间，声波图的纵轴表振幅，频谱图的纵轴表频率 (颜色最深的频率带就是共振峰频率, 默认的频率区间在 0~5000 赫兹)，颜色深浅则表示能量大小。

2. 点击窗口上方 Pitch > Show pitch, 该声音文件的基频曲线 f0 (蓝色) 就会出现在频谱图上。频谱图右边框就会出现默认的基频丈量区间 0~500 赫兹 (蓝色数字)。鼠标指针移动到蓝色曲线上的任意一点，右边就会显示这一点的基频值。见图 5 黑框所示：

3. 点击窗口上方的 Intensity > Show intensity， 该声音所对应的强度曲线 (黄色) 也会出现。鼠标点击曲线上的任意一点，右边框就会显示这一点的音强值 (绿色数字)， 默认区间为 50-100 分贝。见图 5 红框所示：

4. 编辑窗口的左下方有几个按钮作缩放图谱之用，见图 6 红框处。“all” 表示窗口长度则是该声音文件的整个长度。此时，对于长文件来说，细节就看不清了，那么需要放大， 右边的 “in” 就表示 zoom in 放大，每次点击都会在前一次放大的基础上再放大，窗口可能就显示 10 毫秒时间段内的声学信息，甚至更短，意味着细节就更加明显。再右边的 “out” 则表示 zoom out 缩小，有时候过于细节又难以把握全局，特别是对于语调来说，缩小放大一起可以帮你调节到最佳观察范围。 “sel” 表示 select, 必须建立在你已

经从频谱图中选取了一截的基础上，点击它就可以将窗口直接缩放至只包含选取的这一段。选取其中一段声音，只需要将鼠标从开始处点击不放，然后拖拉至结束出放开。选取部分成粉红色。见图 6：

图 5：

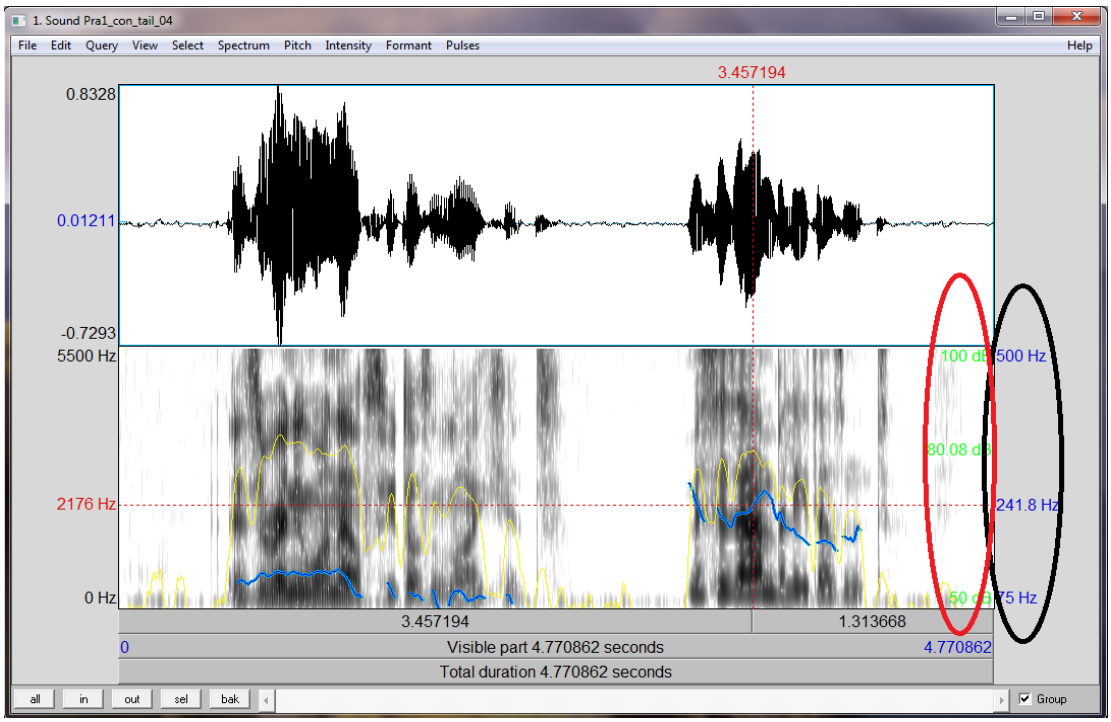
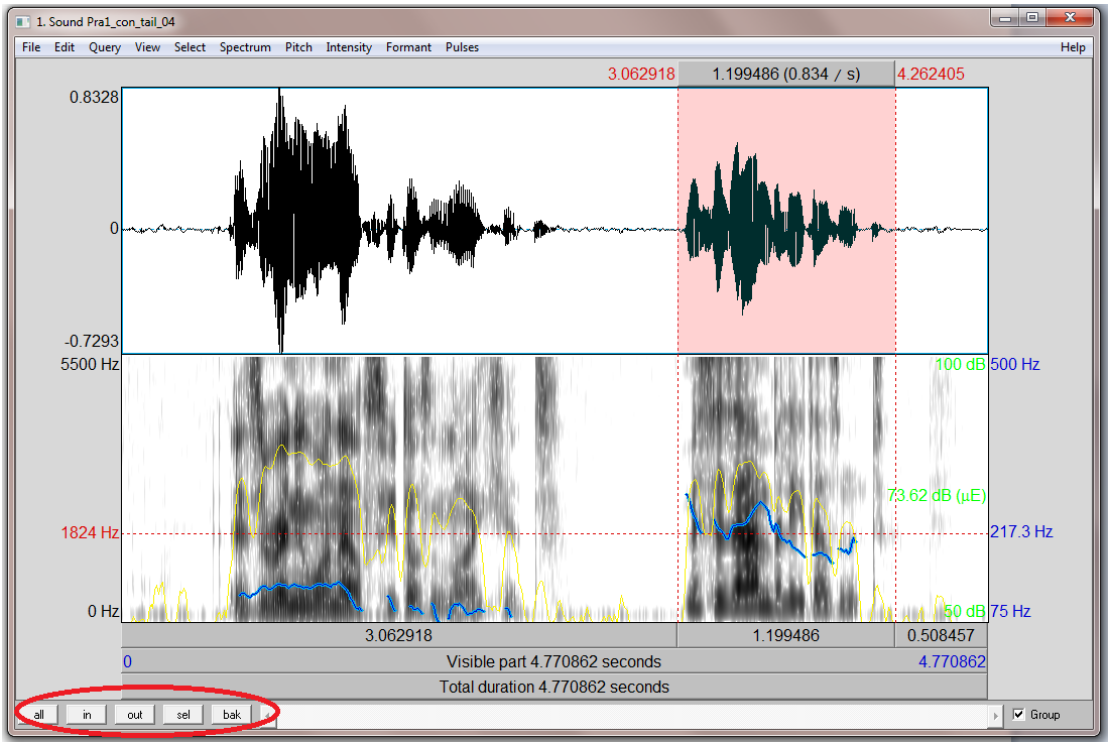


图 6：



5. 窗口下面有三条时间轴，最下面一条是整个声音文件的长度，点击任何一点都可以播放整个声音文件。倒数第二条是目前窗口显示的长度，点击任何一点，播放的就是这个窗口显示的音频部分。 最上面一条时间轴则最灵活。如图 6 所示，被选取的部分在时间轴上出现了相应的两条边界，点击两条边界之间任意一点都可以播放选取的部分。点击左边界前任一点，播放的就是前面那段可视音频，点击右边界后任一点，播放的就是之后那段可视音频。所以，简单来说就是，你想听哪一段就选取那一段，然后点击相应时间轴就可以了。

<u>查看编辑已标注的文件 (Open and edit sound files and their annotations)</u>

1. 从目标窗口中打开一个声音文件 (.wav) 和其相对应的标注文件 (.Textgrid), 这个标注文件的文件名一定是跟它的声音文件相同的，只有扩展名称不一样。选取文件的时候摁住 Ctrl 健不放，将声音文件和标注文件一并选取，然后点确认。

2. 然后两个文件就会出现在目标窗口中，如图 7 所示。此时同样的方法将两个文件同时选取，点击右侧菜单中的 View & Edit (见红色打圈处)，然后就会弹出如图 8 所示的一个新的窗口

图 7：

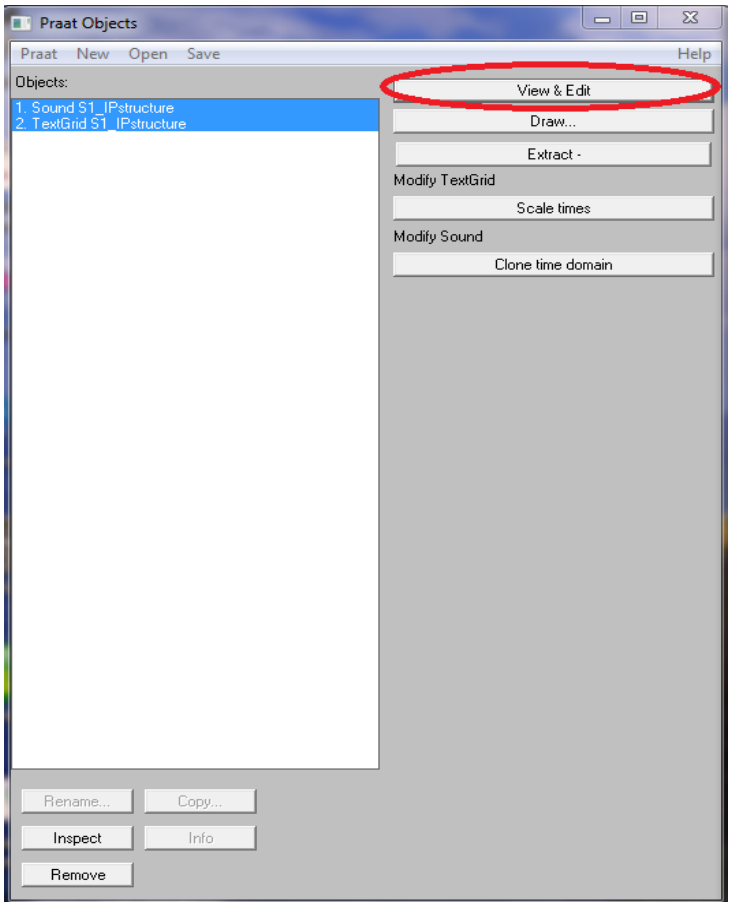
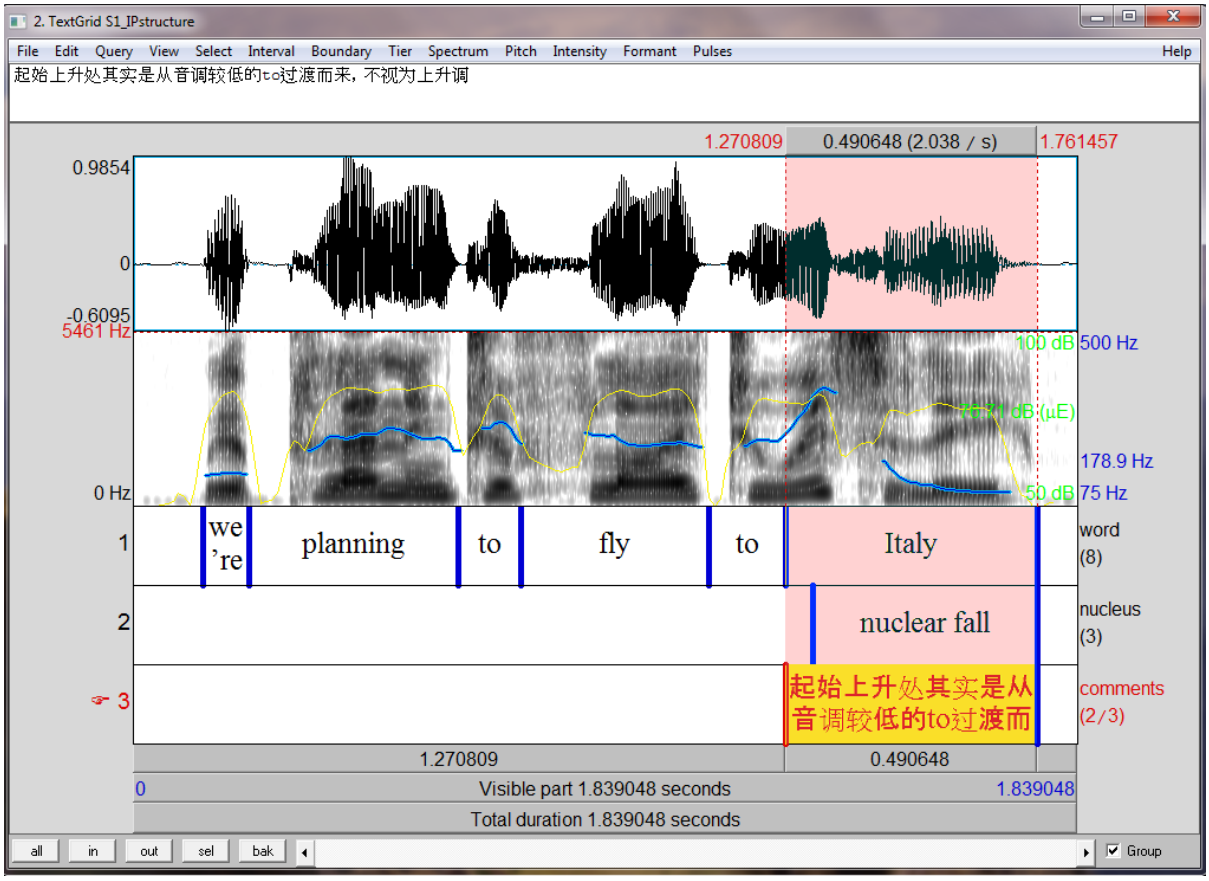

Praat Objects
Praat New Open Save Help
Objects:
1. Sound S1_IPstructure
2. TextGrid S1_IPstructure
View & Edit
Draw...
Extract -
Modify TextGrid
Scale times
Modify Sound
Clone time domain
Rename...
Copy...
Inspect
Info
Remove

图 8：



3. 在之后的语调训练里，所有的声音训练材料都附有相应的标注文件供同学们快速定位目标语调。如图 8 所示，每个声音文件都伴有三行标注，第一行为单词标注，第二行为调核，第三行为附注。选取任何一段标注，点击下面相对应的时间轴，就会听到这段标注的声音。每个训练的重点难点都在第二和第三行标注上，请大家一定注意！通常第三行附注较长，点击它可以在整个窗口上方看到完整的附注。

Question:

从第一行标注上可以看到，整个语流其实是连贯的，那为什么语调曲线不是连贯的，而是有断裂的呢？

准备完毕！

***Notes:** *The terminology bit is adapted from Xiong Ziyu's "Praat 语音软件使用手册". The first three graphs are from Jalal Al-Tamimi's Praat practical handout 1.*

Praat-Session 1

请认真按照步骤，一个任务一个任务，一个步骤一个步骤地完成！

任务１：不同降调的感知和练习

１． 打开声音文件 Task_1.wav 和其标注文件 Task_1.Textgrid

２．请认真听取声音文件，观察语调曲线。因该声音为中老年男性，可以将 pitch 区间设置在７５－４００赫兹，以便更好地观察。

３．四个 now 都用降调读出，但每个在调子长度，调子起伏上都不尽相同。

４．请按照你观察到的这四个降调的不同模仿这四个降调，并录音，存储，观察你自己的降调跟模板是否相似。注意：每个人本身的声调存在生理上的差别，故无须比较你自己的声调绝对值高低是否和模板相同，只需注意调子长度和起伏度是否相似。

５． 如果你对自己的表现不够满意，可以重复录音，再对比，直到满意为止。

任务２：不以辅音结尾的单音节词的降调实现

１． 打开声音文件 Task_２.wav 和其标注文件 Task_２.Textgrid

２．请认真听取声音文件，观察语调曲线。如果频谱图没有出现，则需要 zoom in。因该声音为中年女性，可以将 pitch 区间设置在７５－６００赫兹，以便更好地观察。（至此，以后不再提示，男性区间设置在７５－４００，女性则为７５－６００）

３．这五个单音节词都用降调读出，都以元音结尾。其中 "go" 以浊辅音开始，所以音调曲线一开始就出现；而其余四个都以轻辅音开始，所以音调曲线并没出现直到开始发出元音。

４． 观察每个调子的长度，和起伏度，读出他们并录音。

5． 将你自己的语调曲线与模板对比。

6． 重复步骤 2 – 5 直到你对自己的语调满意为止。

任务 3 ：以浊辅音结尾的单音节词的降调实现

1． 打开声音文件 Task_ 3 .wav 和其标注文件 Task_ 3 .Textgrid

2 ．请认真听取声音文件，观察语调曲线。如果频谱图没有出现，则需要 zoom in。

3 ．这五个单音节词都用降调读出，都以浊辅音结尾。其中 “bad”, “good” 虽然是浊辅音结尾，但都已经清化了，所以没有音调曲线。

4． 观察每个调子的长度，和起伏度，读出他们并录音。

5． 将你自己的语调曲线与模板对比。**注意**：如果你的“bad”, “good”音调曲线在整个单词结尾处结束的话，可能你加了央元音在词尾，念成了 / bædə / 和 / gʊdə /

6． 重复步骤 2 – 5 直到你对自己的语调满意为止。

任务 4 ：以一清辅音结尾的单音节词的降调实现

1． 打开声音文件 Task_ 4 .wav 和其标注文件 Task_ 4 .Textgrid

2 ．请认真听取声音文件，观察语调曲线。如果频谱图没有出现，则需要 zoom in。

3 ．这四个单音节词都用降调读出，都以清辅音结尾。其中 “first”, “six”是两个连续辅音结尾。注意观察这些单音节词的长度，跟任务 3 的词作对比，他们的相似之处在于：

4． 注意观察每个调子的长度，和起伏度，跟任务三的调子作对比，他们的不同之处在于：

5． 朗读这四个单词并将之录下来和模板作对比。**注意**：如果你的 “stop”, “first”, “right” 的音调曲线在结尾清辅音之后还有出现的话，说明你可能加了一个元音在词尾，念成了 / stɔpʊ / 和 / fəstə / / raitə /

6． 重复步骤 5 直到你对自己的语调满意为止。

（3， 4 步骤答案：这四个单词的长度跟任务 3 相似，但他们的音调曲线明显比任务三的短，系末尾以清辅音结束的过，但听感上不造成影响）

任务 5 ：双音节词的降调实现

1． 打开声音文件 Task_ 5 .wav 和其标注文件 Task_ 5 .Textgrid

2 ．请认真听取声音文件，观察语调曲线。如果频谱图没有出现，则需要 zoom in。

3 ．这八个都是词重音在第一个音节的双音节词。注意观察音调曲线，里面出现了两种降调，一种是纯降调，另一种是升降调。留意第三行附言标注，详细解释了这些降调是怎么实现的

4． 模仿这些词的语调实现。将你自己的语调曲线与模板对比。注意升降调和降调的实现是否跟模板相似。绝对音高不必一样，主要注意调幅和相对音节长度。

5． 重复步骤 4 直到你对自己的语调满意为止。

任务 6 ：降调在短句中的应用及语音实现

1． 打开声音文件 Task_ 6 .wav 和其标注文件 Task_ 6 .Textgrid

2 ．请认真听取声音文件，观察语调曲线。Zoom in 来观察详细音调曲线和完整标注及附言。

3 ．这个任务包含 4 个短句对话，每句话的调核都在最后一个单词的重读音节上。标注一共有四层，第一层为对话，点击可听完整对话。第二层为单词，第三层是调核位置及调型，第四层为附言。仔细观察第三和第四层。

4． 观察每句话的调核时长和调核前所有单词的单个时长，其明显差异在于：

5． 仔细观察每句话的调核音高变化和其前面所有单词的音高变化，其明显差异在于：

6． 观察仔细后，模仿这些对话。将你自己的语调曲线与模板对比。注意调核的实现是否跟模板相似，同时也要注意调核前面作为一个整体跟模板的实现是否相似。主要注意调幅和相对音节长度。

7． 重复步骤 6 直到你对自己的语调满意为止。

任务 7 ：不同长度调冠调头和末尾调核的实现

1． 打开声音文件 Task_ 7 .wav 和其标注文件 Task_ 7 .Textgrid

2． 这个任务是包含 incredible 这个单词的五个短句，incredible 在每一句的句尾，充当调核降调，而该单词的词重音在第二个音节-cre-， 所以降调从这个音节开始延伸至最后一个音节。注意观察调核的语音实现。

3． 现在注意观察调核前各个部分的语音实现，尝试数一数他们的音节数，然后对比实现调核降调的三个音节数，它们在时长上有什么相似或不同之处：

4． 对比调核前整个部分的音高变化和调核的音高变化，不同之处在于：

5． 观察仔细后，录下你自己的朗读，对比你和模板之间的异同，哪些相似，哪些不尽相同：

6． 重复 5 直到你和模板的语音实现基本相似

任务８：降调调核及不同长度的调尾的语音实现

１． 打开声音文件 Task_８.wav 和其标注文件 Task_８.Textgrid

２． 这个任务是以 so that(’s) 这两个单词开头的五个短句，that’s 充当调核，其调子为降调，后面跟着不同长度的调尾，通通帮助其将降调实现完整。注意留意第四栏标注附言。

３． 现在注意观察调核后各个部分的语音实现，它们的调子是如何变化的：

４． 观察仔细后，录下你自己的朗读，对比你和模板之间的异同，哪些相似，哪些不尽相同：

５． 重复４直到你和模板的语音实现基本相似。

Praat session 2

♥♥请认真按照步骤，一个任务一个任务，一个步骤一个步骤地完成！

任务 1：升调的语音实现:

1. 下面五句话的调核所在处已经用下滑线标出，请用升调实现每句话的调核，然后朗读出每句话，录音并保存至任务窗口。

a. Who?

b. You want to talk to who?

c. Chicken?

d. All of us?

e. What did you say her name was?

2. 录好音后打开 Task_1.wav 和 Task_1.Textgrid。

3. 确保 pitch setting 的区间在 75-600 赫兹。

4. 仔细观察对比你的录音和模板对调核及调尾升调的实现，填写：

4-1：你和模板调核处的音调及音域差别或相似之处是：

4-2：你和模板调尾处的音调及音域差别或相似之处是：

4-3：你和模板调核前所有部分（句子 b）音调及音域的差别或相似之处是：

5. 按照模板的语音实现模仿这五句话的调核升调直到跟模板相似为止。

任务 2：降升调的语音实现

1： 用降升调念出以下五个单词 (重读音节已被标记出来)并录音保存至任务窗口：

Soon no true today again

2：用降升调念出下面四个单词 （词重音都在首音节，已划出），并录音和保存：

Nearly partly virtually happily

3：用降升调念出下面的单词或短句（同样已用下滑线标出调核音节），录音并保存：

Regrettable reportedly allegedly I think so He said so I hope so

4：现在看你能否总结出以上三组单词或短语彼此间的区别？从调群成分分析

第一组调核位置： 有无调头调尾：

第二组调核位置： 有无调头调尾：

第三组调核位置： 有无调头调尾：

5：现在打开文件 Task_2.wav 和 Task_2.Textgrid，观察以上三组不同调群结构的词汇短句的降升调的语音实现。比较你自己的调型和模板的调型的异同：

第一组你的：

模板的：

第二组你的调核调型：

模板的调核调型：

你的调尾调型：

模板的调尾调型：

第三组你的调核调型：

模板的调核调型：

你的调尾调型：

模板的调尾调型：

6：比较完成后，重新模仿模板的语调实现，看是否你的实现逐渐向模板靠近。如果没有，重复观察并模仿直到满意为止。

任务 3：definitive fall 和 implicational fall-rise 的应用

1：请看下面这则对话，根据情景，在括号中标注出每句话调核的声调（调核见下滑线处）：

A: What can we have for tea? (　　)

B: Well we've got some strawberries. (　　)

A: So what's the problem? (　　)

B: We haven't got any cream. (　　)

2：现在根据你自己的调核声调判断，朗读出这段对话，并录音。

3：打开并观察你自己的录音，语音实现是否跟你的判断一致。是，请进行下一步；否，请重新录音并查看，直到符合你的判断为止。

4：打开文件 Task_3.wav 和 Task_3.Textgrid。观察模板的调核语调是否如你判断。是，那他是怎么实现的。不是，那为什么他会用这些调子？

是：

不是：

5：仔细观察并模仿整个对话，不光是调核和调尾部分，还有调头和调冠。然后重新朗读并录音，直到你自己满意为止。

任务 4：implicational fall-rise 的更多应用

1：观察下面两则对话，打开文件 Task_4.wav 和 Task_4.Textgrid，仔细听 B 的两句回应，在括号内写上他可能省略或者暗示的话：

A: Green and blue are primary colours.

B: Well blue is. （　　）

A: What a lovely voice!

B: Yes, she has a lovely voice. （　　）

2：观察 B 是如何在两句话里实现降升调的。注意，当调核不在句末时，比如第一句，调核的时长和在句末充当调核的时长（第二句）有什么样的差别吗，为什么？

3：模仿 B 的两句话，然后对比观察模板和你自己的实现有何差异。直到满意为止。

任务 5：insistent fall 的用法

1：观察下面两段对话（每句话的调核已经用下滑线标出），请在每句话后面的括号内写出调核语调 （第一段最后一句有两个调群，故有两个调核，分别写出两个调核的调子）

A: Have you come far?（ ）

B: Sorry?（ ）

A: I said, have you come far?（ ）

A: Has Mrs Partington been in?（ ）

B: Sorry?（ ）

A: Has Mrs Partington been in?（ ）

2：现在按照你括号内的调子念出上面这两段对话，并录音。

3：观察你自己的朗读，调核和调尾的实现是否如你自己的答案。

是，那是怎么实现的：

不是，那是哪儿不对：

4：打开文件 Task_5.wav 和 Task_5.Textgrid，观察模板对所有调核的实现是不是如你答案所写？

如果不是，那他是什么调核调子？

5：仔细观察调核，调尾对调核调子的实现，然后观察调头调冠，如果有的话，是怎么实现的。然后模仿并朗读。

6：对比观察模板所有部分和你自己的朗读是否相似（注意：不是完全相同，每个人在语速和基频上都有生理上的差别），如果不是，请重新模仿直到相似为止。

任务 6：升调的用法

1：观察下面三段对话，你能说出 B 的所有问句属于哪种句型吗，它们通常用的语调是什么调子？

A: I want to tell you something.

B: You what? I can't hear you.

A: They've finished the job.

B: Finish the job?

A: She took a tonga.

B: She took a what?

2：打开文件 Task_6.wav 和 Task_6.Textgrid，观察模板是如何实现这类句型的。

3：模仿并录下你自己的朗读。

4：观察这三句 B 问句调核和调核前所有部分跟你自己语音实现的异同，

第一句调冠音域和音高起伏情况：

第一句调核音域和音高起伏情况：

第二句调冠调头音域和音高起伏情况：

第二句调核音域和音高起伏情况：

第三句调冠调头音域和音高起伏情况：

第三句调核音域和音高起伏情况：

5：观察完后，如有需要，请重新录音，直到你自己满意为止。

Praat session 3

♥♥请认真按照步骤，一个任务一个任务，一个步骤一个步骤地完成！

任务 1：无标记的调核置放——单句

1：请划出下面四句话的调核，假设它们都以通常情况出现。

What a disaster!

What are you looking at?

Who was she talking to?

I received a letter from him.

2：现在打开文件 Task_1.wav 和 Task_1.Textgrid, 先听模板的调核是不是如你所想。然后再看蓝色的音高曲线，找出每句话音高开始有明显变化的地方（或音节），然后看那个地方相对应的标注，是否是调核，如果跟你划出来的调核位置不一样，思考一下为什么模板会将调核实现在那儿。

3：听辩和观察调核声调，是降调，升调，还是降升调，为什么？

4：模仿模板的每句话，注意调核的语音实现，它们是怎么跟非调核的地方区分开的。反复模仿直到你跟模板的语音实现相似。

任务 2：无标记的调核置放——对话

1：请划出下面两段对话的调核：

--Who does she work for?

--B and Q.

--What's your number?

--3083.

2：现在打开文件 Task_2.wav 和 Task_2.Textgrid, 先听模板的调核是不是如你所想。然后再看蓝色的音高曲线，找出每句话音高开始有明显变化的地方（或音节），然后看那个地方相对应的标注，是否是调核，如果跟你划出来的调核位置不一样，思考一下为什么模板会将调核实现在那儿。

3：这两段对话里每个调核用的是什么调子？

4：模仿模板每句话，注意调核的语音实现，看它们是如何跟其他非调核的部分区分开来的。用同样的实现方法朗读这两段对话，直到你们的音高曲线相似为止。

任务 3：复合词的重音赏析

1：观察下面几句以复合词（歇体并扫黑）结尾的话，把你认为该念首重音（此处即调核）的音节标注出来：

Is that my *library book*?

I've lost my *credit cards*.

I need some new *running shoes*.

They're in the *departure lounge*.

--Where shall we have our tea? --In the *sitting room*.

Would you like some *Christmas pudding*?

2：打开文件 Task_3.wav 和 Task_3.Textgrid，边听边看模板的调核在哪儿，他的复合词首重音是否跟你的一样？

3：反复观察并模仿这些复合词，直到记熟它们的首重音为止。

任务 4：上/下义词的调核置放：

1：看下列三则对话，划出每句话的调核（每句话只有一个语调调群）：

A: Do you like whist (惠斯特，一种扑克游戏)? B: Oh I like most card games.

A: Will you have some punch (一种冰镇酒)? B: Oh actually I've already got a drink.

A: Do you like ball games? B: Oh I'm quite fond of football.

2：打开文件 Task_4.wav 和 Task_4.Textgrid，边听边看模板的调核在哪儿，他的上/下义词调核置放是否跟你的一样？

答案：A: Do you like whist? （无标记调核，最后一个实意词）

B: Oh I like most card games. (上义词，card games 包含了 whist, 因此属于旧信息，不会成为调核)

A: Will you have some punch? （无标记调核，最后一个实意项）

B: Oh actually I've already got a drink. (上义词，punch 也是 drink 的一种，因此 drink 在此处属于旧信息，不会被赋予调核重音)

A: Do you like ball games? （无标记调核，最后一个复合词重音为调核）

B: Oh I'm quite fond of football. (下义词，足球是球类运动的一种，属于新信息，因此要成为调核)

3：观察并模仿录音，注意调核的语音实现和其他非调核部分的差别，特别是在音高变化幅度和时长上。至于调核调子的使用，看你能不能说出原因？

4：反复模仿，直至每句话的所有成分都跟模板相似为止。

任务 5：新/旧信息的调核置放

1：先看下面几段对话，划出所有语句的调核音节（有两个句子有逗号，表示新调群的出现，所以那两句话各自有两个调核）：

A: Do you all like lasagne (烤宽面条)? B: I do. But I'm not sure whether Barbara does.

A: Do you think we should buy a new sofa? B: If you ask me, it would be a waste of money.

A: He's a famous actor. B: Well not exactly an actor, more a singer.

2：按照你自己的调核置放，用你认为合适的调核调子念出这些对话，然后录音并保存。

3：打开文件 Task_5.wav 和 Task_5.Textgrid，对比下面的答案，看模板的调核及调核调子是否跟你的一样，若不一样，想想他为什么会这么用。

答案：

A: Do you all like lasagne? （无标记调核，最后一个实意项）

B: I do. But I'm not sure whether Barbara does. （implicational fall-rise）

（I 是对比调核，为跟后面的 Barbara 形成对比，所以相应地，Barbara 也是调核）

A: Do you think we should buy a new sofa? (无标记调核)

B: If you ask me, it would be a waste of money. (第一个调核是对比，降升调，意味 my idea might be different than other's，同样表委婉；第二个调核是无标记调核，降调)

A: He's a famous actor. (无标记调核)

B: Well not exactly an actor, more a singer. (对比重复，降升调，委婉; 对比调核，降调)

4：模仿并重复录音，直到跟模板的调核实现一样。

（友情提示：当你觉得调核并不是某句话里面音高变化最大的地方时，仔细观察和听辩模板实现调核和其他地方时有无语速和时长，甚至强度上的差别，这几点也是衡量某个音节是不是调核的标准，以后会讲到。）

任务 6：无标记 vs. 有标记

1：划出下面这句话的无标记调核，并在括号内填上合适的调核调子。

I won’t tell anyone. (　　)

2：同样是这句话，下面有两个有标记的版本，等号后面是它们各自表达的意思，请根据这个意思，划出它们的有标记调核，并在括号内写上调核调子：

I won’t tell anyone. (　　)

= I will tell no one!

I won’t tell anyone. (　　)

=I won’t tell everyone, but just a few.

3：现在根据你自己的判断依次实现这三句话，并录音。

4：打开文件 Task_6.wav 和 Task_6.Textgrid，观察并听辩模板是怎么依次实现它们的，是不是跟你的判断一样？

答案：

I won’t tell anyone. (调核降调，无标记，因 anyone 是代词，不属于实意词。此处没有强调任何人的意思，就是“不会把这件事说出去”。)

I won’t tell anyone. (有标记降调，强调“不会跟任何人说”。)

I won’t tell anyone. (有标记调核，强调“不会跟任何人都讲”。)

5：反复模仿，直到每句话各个部分都跟模板相似为止。

Praat session 4

♥♥请认真按照步骤，一个任务一个任务，一个步骤一个步骤地完成！

任务 1：

1：下面是同一句话用在在三个不同的对话中，请划出每句话的调核（每句话都只有一个调群，故只有一个调核），然后标出每个调核的调子。然后按照你自己的标注朗读这三个情景，并录音。

Dialogue 1: --Do you smoke? --I do. （ ）

Dialogue 2: --I don’t like bacon. --I do. （ ）

Dialogue 3: --Who likes spinach? --I do. （ ）

**Note: spinach 菠菜 [ˋspɪnɪtʃ]*

2：现在检查答案，并打开文件 Task_1.wav 和 Task_1.Textgrid, 听辩及观察模板是不是按照答案所示调核和调核调子来实现的，跟你的判断有出入吗？

Dialogue 1: --Do you smoke? --I do.

（smoke 是无标记调核，降升调，委婉地问；do 是新信息作为调核，降调表肯定）

Dialogue 2: --I don’t like bacon. --I do.

（bacon 是无标记调核，降调；I 是对比调核，降升调表委婉）

Dialogue 3: --Who likes spinach? --I do.

（spinach 是无标记降调；I 是窄焦点，降调表肯定，无保留）

3：现在模仿模板的调核实现，并注意模板是如何通过改变音高和其他方便，比如时长，强度，元音完整度，如果有的话，来跟非调核的其他部分区别开的。反复模仿直到你觉得跟模板方方面面都相似为止。

任务 2：

1：观察下列对话，划出每句话的调核（只有一个），并在括号内解释为什么是它：

--Where’s your passport? --I haven’t got one. （ ）

--Is there a key on the table? --I can’t see one. （ ）

--I’ve been ready for ages. --Why didn’t you say so? （ ）

--Is Jeremy going to get the job? --I don’t think so. （ ）

2：按照你所划出的调核朗读这些对话，并录音。

3：现在打开文件 Task_2.wav 和 Task_2.Textgrid，检查答案，观察并听辩模板是如何分布调核的，是不是跟你的判断一样。

答案：--Where’s your passport? (无标记调核)

--I haven’t got one. (“one” 是 empty pro-form，不能成为调核，除非对比)

--Is there a key on the table? (句末介词短语表地点，不能作调核，除非强调)

--I can’t see one. (“one” 同上)

--I’ve been ready for ages. (无标记调核)

--Why didn’t you say so? (“so” 同样是 pro-form，此处无特殊强调对比，故不充当调核)

--Is Jeremy going to get the job? (无标记调核)

--I don’t think so. (“so” 同上)

4：听辩并观察模板是如何实现这些调核的，有没有用到课上讲的其他几个语音实现的方法除了音调变化外？有，有哪些？

5：模仿这些对话，这次注意非调核的部分的语音实现，该快则快，该短促则短促，有无元音弱读，有无强度变化。反复观察模仿，直到你自己满意为止。

任务 3：动词短语的调核置放

1：观察下面几段包含动词短语的句子和对话，划出每句话的调核：

--You’ve left this line blank.

--Well those details weren’t asked for.

Bring it with you.

Bring your umbrella with you.

Pick the boxes up.

I said, pick them up! （逗号分隔开两个调群，每个调群一个调核）

Now put them down again.

2：现在查看答案是否和你的判断一样。然后打开文件 Task_3.wav 和 Task_3.Textgrid，听辩和观察模板的调核位置。注意第二段话两个 with 做调核和不做调核时的差别，以及第三段话两个 up 在做调核时和不做调核时的语音差别（四个方面）

答案：

--You've left this line blank.

--Well those details weren't asked for.

Bring it with you. Bring your umbrella with you.

Pick the boxes up. I said, pick them up! Now put them down again.

3：现在尽可能地模仿这几段话，直到每个部分（不光是调核）的语音实现都跟模板类似为止。

任务 4：

1：下面三句用斜体表示的话有一个共同的名称，叫做什么？（　　）

Wait and see which way the wind is blowing.

--What's the matter?

--*The baby's crying.*

(phone call) *Hello, this is Jimmy speaking.*

2：划出这三句话的调核，然后根据你自己的调核将这三句话念出来并录音。再检查答案是否正确：（这三句话叫做 event sentences 事件句）

Wait and see which way the wind is blowing.

--What's the matter?

--*The baby's crying.*

(phone call) *Hello, this is Jimmy speaking.*

3：打开文件 Task_4.wav 和 Task_4.Textgrid，听辩和观察模板的调核位置是否如答案所示。观察模板是如何使调核突显出来的，采用的策略有：

4：然后模仿模板直到你认为你们用的策略相同为止。

任务 5：长篇幅的调核置放

1：下面这段话是餐厅服务生在介绍今天的菜单。语调调群已经被 | 切割好了。你要做的就是划出每个调群的调核。（另外，如果时间还有剩余，附加任务是当你确定了调核后，再找出这个调群的其他重音）

Welcome | to Bellamy's Restaurant, Ladies and Gentlemen! | I'm your waiter this evening, | and I'd like to go through the menu with you. | The first course | offers a wide choice of starters. | I'd particularly recommend | the angels on horseback, | the pumpkin soup | or the celery soup. | For the main course | we have steak, lamb or fish, | or also a vegetarian alternative. | I believe the rump steak | is particularly good tonight. |

2：现在打开文件 Task_5.wav 和 Task_5.Textgrid，首先播放全部音频，边听边看你划出来的调核是不是模板都实现了，还有哪些是你没划出来，而模板也实现了的；还有哪些是你划出来，而模板并没实现的，通通作上标记。然后想一想为什么你的调核（和非调核重音）置放跟模板不一样，能找出原因来吗？

3：检查下面的正确答案，下滑线表明调核 （` 表明非调核重音）。

Welcome | to `Bellamy's Restaurant, Ladies and Gentlemen! | `I'm your waiter this evening, | and I'd `like to go through the menu with you. | The first course | offers a `wide choice of starters. | I'd particularly recommend | the `angels on horseback, | the `pumpkin soup | or the celery soup. | For the main course | we have `steak, `lamb or fish, | or `also a vegetarian alternative. | I believe the rump steak | is particularly good tonight. |

4：模仿整段话，然后录下你自己的朗读，也可以一句一句地录，然后观察对比。直到所有的调核和非调核的实现都跟模板相似为止。

Praat Practice 5

请根据每个任务的每一步提示，完成整个练习 ☺

任务 1：调群切分——副词

1：下面四个对话都包含了副词或副词短语，用“/”切分调群如果你认为有必要的话。

A: What are you doing tonight? B: I’ve got a meeting actually.

A: Was the cheese still OK? B: No, it had gone mouldy, naturally.

A: What did you think of the sermon? B: As a matter of fact, it was pretty dire.

A: How did the accident happen? B: Celia, most regrettably, wasn’t paying attention.

*sermon: 布道 dire: 完全不怎么样的

2：下面是参考答案，模板就是按照这些切分来念的四段对话。看你能说出它们为什么要这样切分吗。

A: What are you doing tonight? B: I’ve got a meeting / actually.

A: Was the cheese still OK? B: No, / it had gone mouldy, / naturally.

A: What did you think of the sermon? B: As a matter of fact, / it was pretty dire.

A: How did the accident happen? B: Celia, / most regrettably, / wasn’t paying attention.

3：打开文件 Task_1.wav 和 Task_1.Textgrid, 一边听一边观察模板是用了哪些手段来实现调群切分的 （回顾课上讲的四种方法）。

4：模仿这四段话，听辩并观察你的调群边界是不是跟模板用的策略一样，如果不是，尽量模仿，直到相似为止。

任务 2：

1：观察下面四段对话，看你能否总结出 B 的回答属于哪种句型。然后用“/”划出调群边界如果你觉得有必要的话，并且划出每个调群的调核重读音节。

A: I don’t like Betty’s behaviour. B: It’s her attitude I can’t stand.

A: How have the children been? B: It's Marvin who's been causing all the trouble.

A: What about these dirty marks? B: What granny always did was soak them in vinegar.

A: I love that tree. B: What my neighbour want is for me to cut it down.

2：下面是模板实现的调群边界和每个调群的调核，对比是否跟你的判断一样，若不一样，想想模板为什么会这样划分。

A: I don't like Betty's behaviour. B: It's her attitude / I can't stand.

A: How have the children been? B: It's Marvin / who's been causing all the trouble.

A: What about these dirty marks? B: What granny always did / was soak them in vinegar.

A: I love that tree. B: What my neighbours want / is for me to cut it down.

3：下面打开文件 Task_2.wav 和 Task_2.Textgrid，仔细听辩并观察，每个 B 答句的调群边界由哪些语音指征来完成，在相应的括号内打勾。

第一句：停顿（ ） 起首轻音节（ ）

末尾音节延长（ ） 非重读轻音节音高重设（ ）

第二句：停顿（ ） 起首轻音节（ ）

末尾音节延长（ ） 非重读轻音节音高重设（ ）

第三句：停顿（ ） 起首轻音节（ ）

末尾音节延长（ ） 非重读轻音节音高重设（ ）

第四句：停顿（ ） 起首轻音节（ ）

末尾音节延长（ ） 非重读轻音节音高重设（ ）

4：根据你的判断，仔细模仿模板对调群边界的语音实现，并录音。

5：观察对比你的调群边界实现和模板的差异，反复模仿，直到相似为止。

任务 3：

1：下面是一段长对话，观察并划出每句话的调群边界如果你觉得应该有的话，没有的句子则不用划。然后划出每个调群相应的调核。

A: What's Eve's number?

B: Four six one eight.

A: Sorry?

B: Four six one eight.

A: That’s not a proper number.

B: Well it has four nine one first, of course.

A: So what’s the full number?

B: Give me strength. Four nine one four six one eight.

A: Thank you. You’ve got a problem?

2：检查答案，看是否跟你的判断一样，如果不一样，能说出你自己的版本和模板的划分的理由吗？

A: What’s Eve’s number?

B: Four six one eight.

A: Sorry?

B: Four / six / one / eight.

A: That’s not a proper number.

B: Well it has four nine one first, of course.

A: So what’s the full number?

B: Give me strength. / Four / nine / one / four / six / one / eight.

A: Thank you. / You’ve got a problem?

3：模板的调群切分如上面答案所示，现在打开文件 Task_3.wav 和 Task_3.Textgrid，仔细听辩和观察模板是如何实现调群切分和调核的， 有哪些语音策略是模板运用了的？

4：模仿并录下你自己的朗读。对比观察你对调群切分和调核的语音实现是否和模板相似。如果没有，请反复观察并模仿，直到你自己满意为止

Praat Practice 6

请按照步骤，一一完成！

任务 1；附加疑问句

1：请打开文件 Task_1.wav 和 Task_1.Textgrid, 仔细听下列对话，然后在每个对话后面的括号内填上这句附加疑问句的言外之意。（调群和调核已经标注出来）

--He's Czech.　　--Polish, | isn't he? （　　）

--It's a beautiful day, | isn't it? （　　）

--It looks like rain.　--It does, | doesn't it. （　　）

2：现在模仿这几句附加疑问句，直到跟模板相似为止。

任务 2：Leading tones 的应用

1：观察下面已经被切分好的语句，按照这些调群切分，分别录下你的朗读。

Fortunately, / I was wrong.

Then, / I saw a dog.

Today / we're going to do grammar.

Mrs Ashton / will be taking the children.

As for you, / I'll deal with you later.

On the table / you'll find a jug.

If I were you, / I'd wait and see what happens.

2：每句话的第一个调群由于不是完整的句子成分，它们的调核调子都由表非独立意义的降升调呈现。第二个调群都是独立成分，故都是降调。现在打开文件 Task_2.wav 和 Task_2.Textgrid，将 pitch 区间设置成 75-400Hz, 仔细听辩和观察模板是不是这样念的。然后对比你的朗读，如果你自己不是这样实现的，听听哪种版本更好，是你的呢，还是模板的？

3：现在模仿模板的朗读，按照他的方式念出这几句话。所有调群的调核都是非标记性调核，模仿的时候，注意观察调核的语音实现和非调核部分有何差异。

调核标注：

Fortunately, / I was wrong.

Then, / I saw a dog.

Today / we're going to do grammar.

Mrs Ashton / will be taking the children.

As for you, / I'll deal with you later.

On the table / you'll find a jug.

If I were you, / I'd wait and see what happens.

4：反复观察并模仿，直到所有成分都跟模板一样。

任务 3：Trailing tones 的应用

1：将下列对话的调群和调核划出来，然后用你觉得合适的语调念出来，并录音。

--What can I do for you, sir?

--I'd like this tie, please.

--And for you, madam?

--Some paper, if you'd be so kind.

2：这两个对话的第二句答语的结构都是主句加非完整结构的附加句，而且主句在前，附加句在后，所以主句都由调核降调完成，而末尾附加句都由升调完成。每句话的调群切分和调核置放如下：

--What can I do for you, sir?

--I'd like this tie, / please.

--And for you, madam?

--Some paper, / if you'd be so kind.

3：打开文件 Task_3.wav 和 Task_3.Textgrid，将 pitch 区间重新设置回 75-600 赫兹，仔细听辩观察模板是不是按照答案所示调群切分和调核置放实现的这两个对话。再对比你自己的判断，你觉得是模板的比较好，还是你自己的比较好？

4：现在反复模仿模板的实现，注意调核和非调核的语音实现的差异，及调群边界的语音指征。对比模仿直到你自己的实现跟模板相似为止。

任务 4：调头+调核 语调调型实现

1：下面每句话都只有一个调群。调核（下滑线）和调头首重音（`）都已经标记出来。打开文件 Task_4.wav, 仔细听调核和调头的音调走势，将调型填在括号内。

`Better than ever. （调头：　　调核　　）

It's `nearly ready. （调头：　　调核　　）

`Raring to go? （调头：　　调核　　）

Is `that your partner? （调头：　　调核　　）

`Not at the moment. （调头：　　调核　　）

I'm `awfully sorry. （调头：　　调核　　）

`Don't worry. （调头：　　调核　　）

It `doesn't matter. （调头：　　调核　　）

`Saturday's hopeless. （调头：　　调核　　）

We've `only just begun. （调头：　　调核　　）

2：检查答案，看你能否说出这样处理调核调头语调的原因，从情感投入方面思考。

High level + fall: `Better than ever.

It's `nearly ready.

High level + rise: `Raring to go?

Is `that your partner?

High falling + fall-rise: `Not at the moment.

I'm `awfully sorry.

Low level + rise: `Don't worry.

It `doesn't matter.

Low rising + fall (protest): `Saturday's hopeless.

We've `only just begun.

3：现在打开文件 Task_4.wav 和 Task_4.Textgrid, 对照答案，仔细听辩观察模板的实现。模仿并录音，直到跟模板相似为止。

任务 5：低 vs. 高调冠及情感体现

1：打开文件 Task_5.wav 和 Task_5.Textgrid，一边听一边看下列低/高调冠对比，看你能否感受其蕴含的情感投入差异 （扫黑部分为调冠）

低调冠	vs.	高调冠
You `mustn't worry.		You `mustn't worry.
I `simply don't believe it.		I `simply don't believe it.
She wasn't.		She wasn't.
I will.		I will.
A handbag?		A handbag?

2：模仿低高调冠的语音实现，对比调核和调头音高走势。直到你自己的朗读跟模板类似为止。

任务 6：对话

1：观察下列对话，在你认为合适的地方划分调群，并标记出调核，调头，调冠，然后用你觉得合适的语调将这段对话念出来，并录音。

Hotel guest: Excuse me, where do I get breakfast?

Receptionist: In the Panorama Restaurant, sir.

Hotel guest: Where's that?

Receptionist: Twenty-seventh floor, sir. Use the lift, over there.

Hotel guest: But the lift only goes to the twenty-fourth floor.

Receptionist: Ah. Use lift number five, sir. That one goes to the twenty-seventh floor.

Hotel guest: I see. Thanks.

2：现在打开文件 Task_6.wav 和 Task_6.Textgrid，一边听一边对照答案，看模板是怎么实现这段对话的。

答案：

Hotel guest: Excuse me, / `where do I get breakfast?

Receptionist: In the `Panorama Restaurant, sir.

Hotel guest: `Where's that?

Receptionist: `Twenty-seventh floor, sir. / `Use the lift, / `over there.

Hotel guest: But the `lift only goes to the twenty-fourth floor.

Receptionist: `Ah. `Use lift number five, sir. / That one / `goes to the `twenty-seventh floor.

Hotel guest: I see. / Thanks.

3：对比你自己的朗读，哪些地方跟模板不一样。注意！比较的时候按照下面列举出的顺序来看：

1）调核，调头的位置是否正确；调群切分是否一样？

2）调核的语音实现方式是否一致？

3）调头，调冠的音高走势是否相似？

4）调群边界的语音指征是否相似？

5）调核和非调核是靠那些语音指征区分开来的？

4：比较完后，在按照模板的方式念出这段对话，反复录音直到你自己满意为止。

Appendix VI: Practice materials for the Audacity group

Audacity Practice-Session 1

I. Introduction

What is Audacity?

"Audacity® is free, open source, cross-platform software for recording and editing sounds." It can be used on many operation systems, such as Windows, Mac, etc.

应用比较广泛的录音和编辑软件，其界面简单，直接，易操作。本试验只涉及录音，听音等基本操作。

II. Start-up

重要概念 (Terminology)

１、采样 (Sampling)

由于声音为模拟连续信号，而计算机只能处理数字离散信号，因此要使用计算机来分析和处理声音，就需要经历模数转换过程[Analog to Digital Converter，即 ADC]，即将模拟的连续信号转换为数字离散信号。采样就是按照一定的时间间隔从模拟连续信号提取出一定数量的样本来，其样本值用二进制码 0 和 1 来表示，这些 0 和 1 便构成了数字音频文件，其过程实际上是将模拟音频信号转换成数字离散信号。

２、采样率 (Sampling rate)

采样率表示了每秒对原始信号采样的次数。显然，在一秒中内采样的点越多，获取的信息越丰富，为了复原波形，一次振动中，至少得有 2 个点的采样，要想使采集到的信号不失真，采样频率规定至少为语音频率的 2 倍，因此要得到一个频率为 10000 赫兹(Hz)的声音，则其采样率至少得大于 20000 赫兹。采样频率越高，数字信号的保真度越高，但同时占用的存储空间也越大。如果采样率低于高频成分频率的两倍，则会产生低频失真、信号混淆现象。

３、采样精度 (Sampling size)

采样精度就是指存放一个采样值所使用的比特数目。当用 8 个比特 (bit)（采样精度为 8 位）存放一个采样值时，对声音振幅的分辨等级理论上为 256 个，即 0 至 255；当用 16

个比特（采样精度为 16 位）存放一个采样值时，对声音振幅的分辨等级理论上为65536 个，即 0 至 65535。如果您将采样精度设置为 16 位，计算机纪录的采样值范围则为-32768 至 32767 之间的整数。采样率和采样精度的值越大，记录的波形更接近原始信号，但同时占用的存储空间也越大。

４、声道 (Sound channel)

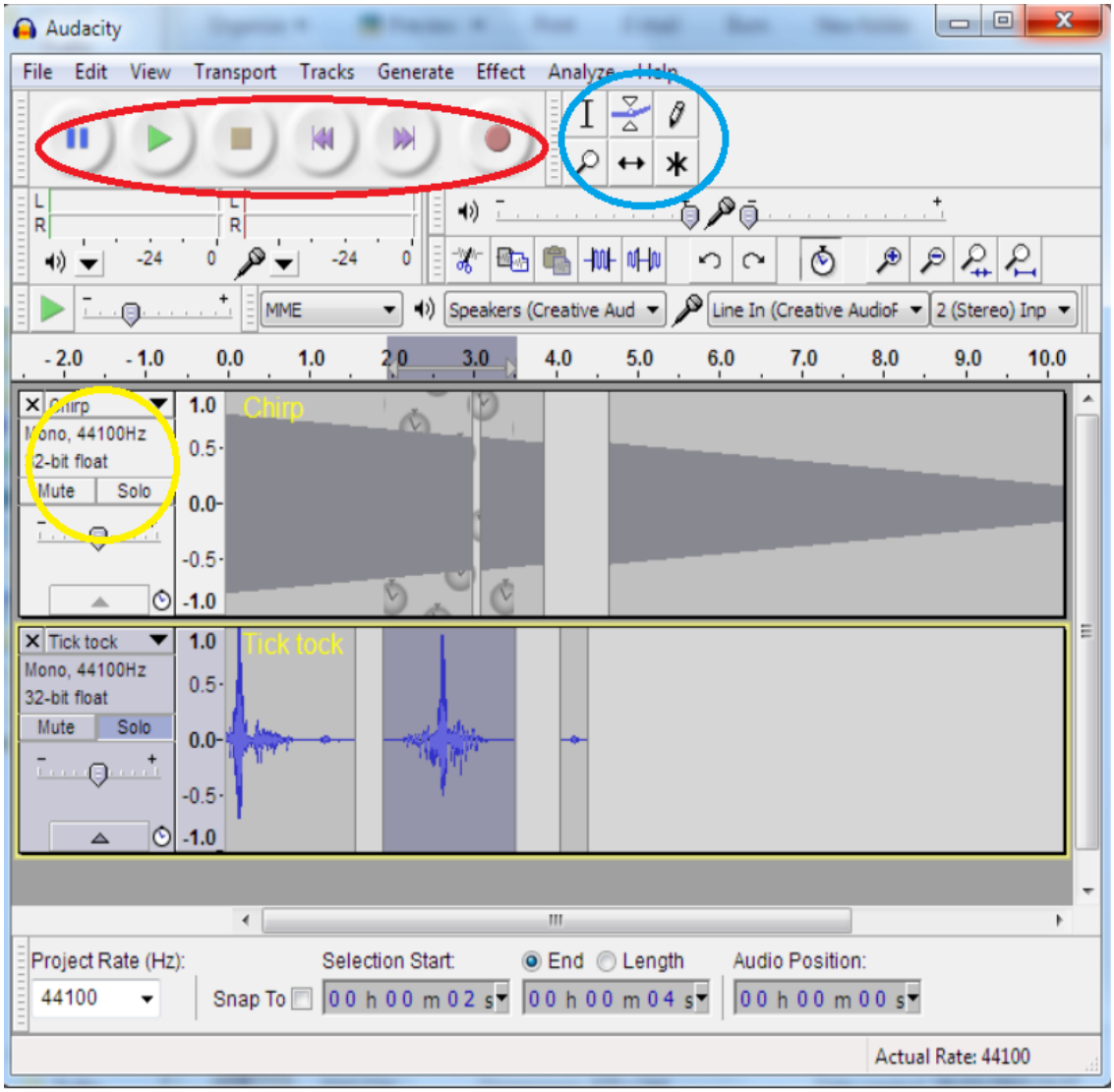
声道指输入或输出信号的通道。通常用多声道来输入或输出不同的信号。如果只需录制一个位置的一种信号时，只要使用单声道就可以了。

声音采集 (Recording a sound file)

1. 登陆你自己的学校账户。离开教室前，记得登出 (log off)。

2. 点击 home 键，在搜索栏里输入 audacity，然后就会出现该软件应用。点击它打开audacity。系统会提示正在安装。等候片刻，安装好后会自动弹出其应用界面，如图 1 所示：

图 1：



红色区域按钮从左到右分别为：暂停播放，播放，停止，上一个音频，下一个音频，录音。录音的默认采样率为 44100 赫兹，默认精度为 32bit, 保持这两个默认值不变，见上图黄色区域。

3. 戴上耳机，将话筒置于你嘴巴左前方 5 厘米处。然后点击录音键，试着说一段话。然后点击停止键。这段话的波形图就会出现在界面中间。默认输入是双声道，两个声道的波形图完全一样。点击播放就可以听你自己的录音了。每次点击录音键然后停止，新产生的音频不会覆盖上一次的录音。也就是说你所有录音都会出现在界面上，除非你点击每个录音波形图最左上方的 x 关闭该音频。

Audacity-Session 1

♥♥ 请认真按照每个步骤，一个任务一个任务地完成！

任务 1：不同降调的感知和练习

1：从 Audacity 中打开模板文件 Task_1.wav，里面用不同的降调读出单音节词 now。但每个 now 在调子长度，调子起伏度上都不尽相同。请认真听录音，从调子长度和起伏度上来讲，这四个降调的区别是:

第一个 now:

第二个 now:

第三个 now:

第四个 now:

2：现在模仿模板里面的四个降调，并用 Audacity 录下来。从调子的长度和起伏度上来讲，对比你自己的实现和模板的相似和不同之处（注意！每个人本身的声调存在生理上的差别，故无须比较你自己的声调绝对值高低是否和模板相同，只需注意调子长度和起伏度是否相似。）

3：如果你自己的实现跟模板始终有差异的话，请反复练习直到你满意为止。

任务 2：不同音节结构的单音节词的降调实现

1：打开模板文件 Task_2.wav，里面用降调朗读了下面三组单音节词，但第一行以元音结尾，第二行以浊辅音 （声带震动）结尾，第三行以清爆音或摩擦音（无声带震动）结尾。请仔细听模板的降调是怎么在它们身上实现的。

Three go four stay sure ( )

Bad nine good sing come ( )

Nice stop first right six ( )

2：对比这三组降调在调子长度和调子起伏度上的差异，在每组单词后面的括号内填上这些差异。

3：现在录下你自己对这三组单词降调的实现，然后反复聆听对比模板，直到你自己满意为止。

任务 3：双音节词的降调实现

1：下面这些单词全部都是双音节词，其词重音都在第一个音节（见下滑线），请打开模板文件 Task_3.wav，仔细听录音是怎么在这些双音节词上实现降调的。

Super!　　Crazy!　　Never!　　Splendid!

Heavens!　Rubbish!　Nonsense!　Awesome!

2：比较 super, nonsense, awesome 和其他单词在实现降调的时候有什么听感上的差别？

Super, nonsense, awesome:

其他：

（提示：这三个单词的第二个音节都是以清辅音开始，意味着声带在发这些音的时候是没有震动的。）

3：模仿这些降调，对比你和模板在语音实现上的异同，然后重复录音，对比，直到你自己满意为止。

任务 4：降调在短语中的应用及语音实现

1：下面是四个短对话，每句话的调核都在最后一个单词的词重音上（被下滑线画出）。请打开文件 Task_4.wav, 仔细听这些对话，特别注意调核的语音实现和前面所有其他单词的语音实现的差异，完成下列空格。

1. --I'll be there by five.

--Great!

2. --The sausage's got burnt.

--What a pity!

3. --We've just got engaged.

--How marvelous!

4. --She's had a baby boy.

--But that’s wonderful!

仔细听辩每句话的调核时长和调核前所有单词的单个时长，其明显差异在于：

调核：（ ）

其他：（ ）

仔细听辩每句话的调核音高变化和其前面所有单词的音高变化，其明显差异在于：

调核：（ ）

其他：（ ）

2：模仿这些对话，特别注意调核音节和其他非调核音节在语音实现上的差异。反复聆听，模仿，对比，直到你自己满意为止。

任务 5：不同长度调冠调头和末尾调核的实现

1：打开文件 Task_5.wav, 里面是包含 incredible 这个单词的五个短句。incredible 在每一句的句尾，充当调核降调，而该单词的词重音在第二个音节-cre-， 所以降调从这个音节开始延伸至最后一个音节。仔细听辩英国人是怎么实现调核和调核前后的所有部分的，并完成下列问题。

Incredible!

How incredible!

That’s incredible!

How utterly incredible!

They’re going to find it utterly incredible!

仔细听录音，用 ` 标出每句话的调头起始音节如果有的话。

调核的时长和非调核其他音节的时长差异是（忽略非调核重音如果有的话）：

（ ）

调核的音高起伏度和非调核其他音节的音高起伏度的差异是（忽略非调核重音的起伏度）：

（ ）

2：按照以上你所听到的差异，模仿这几句话，注意调核的语音实现。反复模仿听辩直到跟模板相似为止。

任务 6：降调调核及不同长度的调尾的语音实现

1：下面是以 so that('s) 这两个单词开头的五个短句，that's 充当调核，其调子为降调，后面跟着不同长度的调尾，通通帮助其将降调实现完整。请打开文件 Task_6.wav 并认真听辩模板，完成下面空格。

So that's it.

So that was the trouble.

So that's what he wants.

So that's what he was getting at.

So that's what he told you yesterday.

调核和调尾在音高特征上的差别是（ ）

调核和调尾在强度上的差别是：（ ）

2：模仿并录下你对这些短句的实现，特别注意你自己的调核调尾的差异性是否和模板一样突出。

3：反复模仿直到你自己满意为止。

Audacity session 2

♥♥请认真按照步骤，一个任务一个任务，一个步骤一个步骤地完成！

任务 1：升调的语音实现:

1. 下面五句话的调核所在处已经用下滑线标出，请用升调实现每句话的调核，然后朗读出每句话，录音并保存。

a. Who?

b. You want to talk to who?

c. Chicken?

d. All of us?

e. What did you say her name was?

2. 录好音后打开 Task_1.wav。

3. 仔细听辩对比你的录音和模板对调核及调尾升调的实现，填写：

3-1：你和模板调核处的音调及音域差别或相似之处是：

3-2：你和模板调尾处的音调及音域差别或相似之处是：

3-3：你和模板调核前所有部分（句子 b）的音调及音域差别是：

4. 按照模板的语音实现模仿这五句话的调核升调直到跟模板相似为止。

任务 2：降升调的语音实现

1： 用降升调念出以下五个单词 (重读音节已被标记出来)并录音和保存：

Soon no true today again

2：用降升调念出下面四个单词 （词重音都在首音节，已划出），并录音和保存：

Nearly partly virtually happily

3：用降升调念出下面的单词或短句（同样已用下滑线标出调核音节），录音并保存：

Regrettable reportedly allegedly I think so He said so I hope so

4：现在看你能否总结出以上三组单词或短语彼此间的区别？从调群成分分析，

第一组调核位置： 有无调头调尾：

第二组调核位置： 有无调头调尾：

第三组调核位置： 有无调头调尾：

5：现在打开文件 Task_2.wav，听辩以上三组不同调群结构的词汇短句的降升调的语音实现。比较你自己的调型和模板的调型的异同：

第一组你的：

模板的：

第二组你的调核调型：

模板的调核调型：

你的调尾调型：

模板的调尾调型：

第三组你的调核调型：

模板的调核调型：

你的调尾调型：

模板的调尾调型：

6：比较完成后，重新模仿模板的语调实现，看是否你的实现逐渐向模板靠近。如果没有，重复听辩并模仿直到满意为止。

任务 3：definitive fall 和 implicational fall-rise 的应用

1：请看下面这则对话，根据情景，在括号中标注出每句话调核的声调（调核见下滑线处）：

A: What can we have for tea? (　　)

B: Well we've got some strawberries. (　　)

A: So what's the problem? (　　)

B: We haven't got any cream. (　　)

2：现在根据你自己的调核声调判断，朗读出这段对话，并录音。

3：打开并听辩你自己的录音，语音实现是否跟你的判断一致。是，请进行下一步；否，请重新录音并查看，直到符合你的判断为止。

4：打开文件 Task_3.wav。听辩模板的调核语调是否如你判断。是，那他是怎么实现的。不是，那为什么他会用这些调子？

是：

不是：

5：仔细听辩并模仿整个对话，不光是调核和调尾部分，还有调头和调冠。然后重新朗读并录音及对比，直到你自己满意为止。

任务 4：implicational fall-rise 的更多应用

1：观察下面两则对话，打开文件 Task_4.wav，仔细听 B 的两句回应，在括号内写上他可能省略或者暗示的话：

A: Green and blue are primary colours.

B: Well <u>blue</u> is. （ ）

A: What a lovely voice!

B: Yes, she has a lovely <u>voice</u>. （ ）

2：仔细听 B 是如何在两句话里实现降升调的。注意，当调核不在句末时，比如第一句，调核的时长和在句末充当调核的时长（第二句）有什么样的差别吗，为什么？

3：模仿 B 的两句话，然后对比听辩模板和你自己的实现有何差异。直到满意为止。

任务 5：insistent fall 的用法

1：观察下面两段对话（每句话的调核已经用下滑线标出），请在每句话后面的括号内写出调核语调 （第一段最后一句有两个调群，故有两个调核，分别写出两个调核的调子）

A: Have you come <u>far</u>?（ ）

B: Sorry?（ ）

A: I said, have you come far?（ ）

A: Has Mrs Partington been in?（ ）

B: Sorry?（ ）

A: Has Mrs Partington been in?（ ）

2：现在按照你括号内的调子念出上面这两段对话，并录音。

3：观察你自己的朗读，调核和调尾的实现是否如你自己的答案。

是，那是怎么实现的：

不是，那是哪儿不对：

4：打开文件 Task_5.wav，仔细听模板对所有调核的实现是不是如你答案所写？

如果不是，那他是什么调核调子？

5：仔细听辩模板调核，调尾对调核调子的实现，然后听辩调头调冠，如果有的话，是怎么实现的。然后模仿并朗读。

6：对比听辩模板的所有部分和你自己的朗读是否相似（注意：不是完全相同，每个人在语速和基频上都有生理上的差别），如果不是，请重新模仿直到相似为止。

任务 6：升调的用法

1：观察下面三段对话，你能说出 B 的所有问句属于哪种句型吗，它们通常用的语调是什么调子？

A: I want to tell you something.

B: You what? I can’t hear you.

A: They’ve finished the job.

B: Finish the job?

A: She took a tonga.

B: She took a what?

2：打开文件 Task_6.wav，仔细听辩模板是如何实现这类句型的。

3：模仿并录下你自己的朗读。

4：对比听辩这三句 B 问句调核和调核前所有部分跟你自己语音实现的异同，

第一句调冠音域和音高起伏情况：

第一句调核音域和音高起伏情况：

第二句调冠调头音域和音高起伏情况：

第二句调核音域和音高起伏情况：

第三句调冠调头音域和音高起伏情况：

第三句调核音域和音高起伏情况：

5：听辩完后，如有需要，请重新录音，直到你自己满意为止。

Audacity session 3

♥♥请认真按照步骤，一个任务一个任务，一个步骤一个步骤地完成！

任务 1：无标记的调核置放

1：请划出下面四句话的调核，假设它们都以通常情况出现。

What a disaster!

What are you looking at?

Who was she talking to?

I received a letter from him.

2：现在从 Audacity 中打开文件 Task_1.wav, 先听模板的调核是不是如你所想，注意每句话音高开始有明显变化的地方（或音节），那儿就是调核。然后检查答案：

What a disaster!

What are you looking at?

Who was she talking to?

I received a letter from him.

3：听辩调核声调，是降调，升调，还是降升调，为什么？

4：模仿模板的每句话，注意调核的语音实现，它们是怎么跟非调核的地方区分开的。反复模仿直到你跟模板的语音实现相似。

任务 2：无标记的调核置放——对话

1：请划出下面两段对话的调核：

--Who does she work for?

--B and Q.

--What's your number?

--3083.

2：现在在 Audacity 里面打开文件 Task_2.wav, 先听模板的调核是不是如你所想（提示：每句话音高开始有明显变化的地方（或音节）就是调核），如果跟你划出来的调核位置不一样，思考一下为什么模板会将调核实现在那儿。

答案：

--Who does she work for?

--B and Q.

--What's your number?

--3083.

3：这两段对话里每个调核用的是什么调子？（答案：全是降调，最后一句 3083 的调核最后有一点上扬，可以认为是降升调）

4：模仿模板每句话，注意调核的语音实现，听它们是如何跟其他非调核的部分区分开来的。用同样的实现方法朗读这两段对话，直到从听感上你觉得它们相似为止。

任务 3：复合词的重音赏析

1：观察下面几句以复合词（歇体并扫黑）结尾的话，把你认为该念首重音（此处即调核）的音节标注出来：

Is that my *library book*?

I've lost my *credit card*.

I need some new *running shoes*.

They're in the *departure lounge*.

--Where shall we have our tea? --In the *sitting room*.

Would you like some *Christmas pudding*?

2：打开文件 Task_3.wav，边听边想模板的调核是不是在你自己标注的音节。

3：现在检查答案：

Is that my *library book*?

I've lost my *credit cards*.

I need some new *running shoes*.

They're in the *departure lounge*.

--Where shall we have our tea? --In the *sitting room*.

Would you like some *Christmas pudding*?

4：反复观察并模仿这些复合词，直到记熟它们的首重音为止。

任务 4：上/下义词的调核置放：

1：看下列三则对话，划出每句话的调核（每句话只有一个语调调群）：

A: Do you like whist (惠斯特，一种扑克游戏)? B: Oh I like most card games.

A: Will you have some punch (一种冰镇果酒)? B: Oh actually I've already got a drink.

A: Do you like ball games? B: Oh I'm quite fond of football.

2：打开文件 Task_4.wav，边听边判断模板的调核在哪儿，他的上/下义词调核置放是否跟你的一样？

答案：A: Do you like whist? （无标记调核，最后一个实意词）

B: Oh I like most card games. (上义词，card games 包含了 whist, 因此属于旧信息，不会成为调核)

A: Will you have some punch? （无标记调核，最后一个实意项）

B: Oh actually I've already got a drink. (上义词，punch 也是 drink 的一种，因此 drink 在此处属于旧信息，不会被赋予调核重音)

A: Do you like ball games? （无标记调核，最后一个复合词重音为调核）

B: Oh I'm quite fond of football. (下义词，足球是球类运动的一种，属于新信息，因此要成为调核)

3：仔细听并模仿录音，注意调核的语音实现和其他非调核部分的差别，特别是在音高变化幅度和时长上。至于调核调子类型的使用，看你能不能说出原因？

4：反复模仿，直至你觉得听上去每句话的所有成分都跟模板相似为止。

任务 5：新/旧信息的调核置放

1：先看下面几段对话，划出所有语句的调核音节（有两个句子有逗号，表示新调群的出现，所以那两句话各自有两个调核）：

A: Do you all like lasagne (烤宽面条)?　　B: I do. But I'm not sure whether Barbara does.

A: Do you think we should buy a new sofa?　B: If you ask me, it would be a waste of money.

A: He's a famous actor.　　B: Well not exactly an actor, more a singer.

2：按照你自己的调核置放，用你认为合适的调核调子念出这些对话，然后录音并保存。

3：打开文件 Task_5.wav，对比下面的答案，仔细听模板的调核及调核调子是否跟你的一样，若不一样，想想他为什么会这么用。

答案：

A: Do you all like lasagne? （无标记调核，最后一个实意项）

B: I do. But I'm not sure whether Barbara does. （implicational fall-rise）

（I 是对比调核，为跟后面的 Barbara 形成对比，所以，相应地，Barbara 也是调核）

A: Do you think we should buy a new sofa? (无标记调核)

B: If you ask me, it would be a waste of money. (第一个调核是对比，降升调，意味 my idea might be different than other's，同样表委婉；第二个调核是无标记调核，降调)

A: He’s a famous actor. (无标记调核)

B: Well not exactly an actor, more a singer. (对比重复，降升调，委婉; 对比调核，降调)

4：模仿并重复录音，直到跟模板的调核实现一样。

（友情提示：当你觉得调核并不是某句话里面音高变化最大的地方时，仔细听辩模板实现调核和其他地方时有无语速和时长，甚至强度上的差别，这几点也是衡量某个音节是不是调核的标准，以后会讲到。）

任务 6：无标记 vs. 有标记

1：划出下面这句话的无标记调核，并在括号内填上合适的调核调子。

I won’t tell anyone. (　　)

2：同样是这句话，下面有两个有标记的版本，等号后面是它们各自表达的意思，请根据这个意思，划出它们的有标记调核，并在括号内写上调核调子：

I won’t tell anyone. (　　)

= I will tell no one!

I won’t tell anyone. (　　)

=I won’t tell everyone, but just a few.

3：现在根据你自己的判断依次实现这三句话，并录音。

4：打开文件 Task_6.wav，听辩模板是怎么依次实现它们的，是不是跟你的判断一样？

答案：

I won’t tell anyone. (调核降调，无标记，因 anyone 是代词，不属于实意词。此处没有强调任何人的意思，就是“不会把这件事说出去”。)

I won’t tell anyone. (有标记降调，强调“不会跟任何人说”。)

I won’t tell anyone. (有标记调核，强调“不会跟任何人都讲”。)

5：反复模仿，直到每句话各个部分都跟模板相似为止。

Audacity session 4

♥♥请认真按照步骤，一个任务一个任务，一个步骤一个步骤地完成！

任务 1：

1：下面是同一句话用在在三个不同的对话中，请划出每句话的调核（每句话都只有一个调群，故只有一个调核），然后标出每个调核的调子。然后按照你自己的标注朗读这三个情景，并录音。

Dialogue 1: --Do you smoke? --I do. （ ）

Dialogue 2: --I don't like bacon. --I do. （ ）

Dialogue 3: --Who likes spinach? --I do. （ ）

**Note: spinach 菠菜 [`spɪnɪt∫]*

2：现在检查答案，并从 Audacity 中打开文件 Task_1.wav, 听辩模板是不是按照答案所示调核和调核调子来实现的，跟你的判断有出入吗？

Dialogue 1: --Do you smoke? --I do.

（smoke 是无标记调核，降升调，委婉地问；do 是新信息作为调核，降调表肯定）

Dialogue 2: --I don't like bacon. --I do.

（bacon 是无标记调核，降（升）调；I 是对比调核，降升调表委婉）

Dialogue 3: --Who likes spinach? --I do.

（spinach 是无标记降调；I 是窄焦点，降调表肯定，无保留）

3：现在模仿模板的调核实现，并注意模板是如何通过改变音高和其他方便，比如时长，强度，元音完整度，如果有的话，来跟非调核的其他部分区别开的。反复模仿直到你觉得跟模板方方面面都相似为止。

任务 2：

1：观察下列对话，划出每句话的调核（只有一个），并在括号内解释为什么是它：

--Where's your passport? --I haven't got one. （ ）

--Is there a key on the table? --I can't see one. （ ）

--I've been ready for ages. --Why didn't you say so? （ ）

--Is Jeremy going to get the job? --I don't think so. （ ）

2：按照你所划出的调核朗读这些对话，并录音。

3：现在打开文件 Task_2.wav，检查答案，听辩模板是如何分布调核的，是不是跟你的判断一样。

答案：--Where's your passport? (无标记调核)

--I haven't got one. ("one" 是 empty pro-form，不能成为调核，除非对比)

--Is there a key on the table? (句末介词短语表地点，不能作调核，除非强调)

--I can't see one. ("one" 同上)

--I've been ready for ages. (无标记调核)

--Why didn't you say so? ("so" 同样是 pro-form，此处无特殊强调对比，故不充当调核)

--Is Jeremy going to get the job? (无标记调核)

--I don't think so. ("so" 同上)

4：听辩模板是如何实现这些调核的，有没有用到课上讲的其他几个语音实现的方法除了音调变化外？有，有哪些？

5：模仿这些对话，这次注意非调核部分的语音实现，该快则快，该短促则短促，有无元音弱读，有无强度变化。反复模仿，直到你自己满意为止。

任务 3：动词短语的调核置放

1：观察下面几段包含动词短语的句子和对话，划出每句话的调核：

--You’ve left this line blank.

--Well those details weren’t asked for.

Bring it with you.

Bring your umbrella with you.

Pick the boxes up.

I said, pick them up! （逗号分隔开两个调群，每个调群一个调核）

Now put them down again.

2：现在查看答案是否和你的判断一样。然后打开文件 Task_3.wav，听辩模板的调核位置。注意第二段话两个 with 做调核和不做调核时的差别，以及第三段话两个 up 在做调核时和不做调核时的语音差别（四个方面）。

答案：

--You’ve left this line blank.

--Well those details weren’t asked for.

Bring it with you. Bring your umbrella with you.

Pick the boxes up. I said, pick them up! Now put them down again.

3：现在尽可能地模仿这几段话，直到每个部分（不光是调核）的语音实现都跟模板类似为止。

任务 4：

1：下面三句用斜体表示的话有一个共同的名称，叫做什么？（ ）

Wait and see which way the wind is blowing.

--What’s the matter?

--*The baby’s crying.*

(phone call) *Hello, this is Jimmy speaking.*

2：划出这三句话的调核，然后根据你自己的调核将这三句话念出来并录音。再检查答案是否正确：（这三句话叫做 event sentences 事件句）

Wait and see which way the wind is blowing.

--What's the matter?

--*The baby's crying.*

(phone call) *Hello, this is Jimmy speaking.*

3：打开文件 Task_4.wav，听辩模板的调核位置是否如答案所示。观察模板是如何使调核突显出来的，采用的策略有：

4：然后模仿模板直到你认为你们用的策略相同为止。

任务 5：长篇幅的调核置放

1：下面这段话是餐厅服务生在介绍今天的菜单。语调调群已经被 | 切割好了。你要做的就是划出每个调群的调核。（另外，如果时间还有剩余，附加任务是当你确定了调核后，再找出这个调群的其他重音）

Welcome | to Bellamy's Restaurant, Ladies and Gentlemen! | I'm your waiter this evening, | and I'd like to go through the menu with you. | The first course | offers a wide choice of starters. | I'd particularly recommend | the angels on horseback, | the pumpkin soup | or the celery soup. | For the main course | we have steak, lamb or fish, | or also a vegetarian alternative. | I believe the rump steak | is particularly good tonight. |

2：现在打开文件 Task_5.wav，首先播放全部音频，边听边看你划出来的调核是不是模板都实现了，还有哪些是你没划出来，而模板也实现了的；还有哪些是你划出来，而

模板并没实现的，通通作上标记。然后想一想为什么你的调核（和非调核重音）置放跟模板不一样，能找出原因来吗？

3：检查下面的正确答案，下滑线表明调核 （` 表明非调核重音）。

Welcome | to `Bellamy's Restaurant, Ladies and Gentlemen! | `I'm your waiter this evening, | and I'd `like to go through the menu with you. | The first course | offers a `wide choice of starters. | I'd particularly recommend | the `angels on horseback, | the `pumpkin soup | or the celery soup. | For the main course | we have `steak, `lamb or fish, | or `also a vegetarian alternative. | I believe the rump steak | is particularly good tonight. |

4：模仿整段话，然后录下你自己的朗读，也可以一句一句地录，然后听辩对比。直到所有的调核和非调核的实现都跟模板相似为止。

Audacity Practice 5

请根据每个任务的每一步提示，完成整个练习 ☺

任务 1：调群切分——副词

1：下面四个对话都包含了副词或副词短语，用"/"切分调群如果你认为有必要的话。

A: What are you doing tonight? B: I've got a meeting actually.

A: Was the cheese still OK? B: No, it had gone mouldy, naturally.

A: What did you think of the sermon? B: As a matter of fact, it was pretty dire.

A: How did the accident happen? B: Celia, most regrettably, wasn't paying attention.

*sermon: 布道 dire: 完全不怎么样的

2：下面是参考答案，模板就是按照这些切分来念的四段对话。看你能说出它们为什么要这样切分吗。

A: What are you doing tonight? B: I've got a meeting / actually.

A: Was the cheese still OK? B: No, / it had gone mouldy, / naturally.

A: What did you think of the sermon? B: As a matter of fact, / it was pretty dire.

A: How did the accident happen?　　B: Celia, / most regrettably, / wasn't paying attention.

3：打开文件 Task_1.wav, 一边听一边留意模板是用了哪些手段来实现调群切分的 （回顾课上讲的四种方法）。

4：模仿这四段话，仔细听辩你的调群边界是不是跟模板用的策略一样，如果不是，尽量模仿，直到相似为止。

任务 2：

1：观察下面四段对话，看你能否总结出 B 的回答属于哪种句型。然后用"/"划出调群边界如果你觉得有必要的话，并且划出每个调群的调核重读音节。

A: I don't like Betty's behaviour.　B: It's her attitude I can't stand.

A: How have the children been?　B: It's Marvin who's been causing all the trouble.

A: What about these dirty marks?　B: What granny always did was soak them in vinegar.

A: I love that tree.　　B: What my neighbour want is for me to cut it down.

2：下面是模板实现的分裂/假拟分裂句的调群边界和每个调群的调核，对比是否跟你的判断一样，若不一样，想想模板为什么会这样划分。

A: I don't like Betty's behaviour.　B: It's her attitude / I can't stand.

A: How have the children been?　B: It's Marvin / who's been causing all the trouble.

A: What about these dirty marks?　B: What granny always did / was soak them in vinegar.

A: I love that tree.　　B: What my neighbours want / is for me to cut it down.

3：下面打开文件 Task_2.wav，仔细听辩每个 B 答句的调群边界由哪些语音指征来完成，在相应的括号内打勾。

第一句：停顿（　　）　起首轻音节（　　）

　　末尾音节延长（　　）　非重读轻音节音高重设（ ）

第二句：停顿（　　）　起首轻音节（　　）

　　末尾音节延长（　　）　非重读轻音节音高重设（ ）

第三句：停顿（　　）　起首轻音节（　　）

末尾音节延长（　　）　　非重读轻音节音高重设（ ）

第四句：停顿（　　）　　起首轻音节（　　）

末尾音节延长（　　）　　非重读轻音节音高重设（ ）

4：根据你的判断，仔细模仿模板对调群边界的语音实现，并录音。

5：听辩对比你的调群边界实现和模板的差异，反复模仿，直到相似为止。

任务 3：

1：下面是一段长对话，观察并划出每句话的调群边界如果你觉得应该有的话，没有的句子则不用划。然后划出每个调群相应的调核。

A: What’s Eve’s number?

B: Four six one eight.

A: Sorry?

B: Four six one eight.

A: That’s not a proper number.

B: Well it has four nine one first, of course.

A: So what’s the full number?

B: Give me strength. Four nine one four six one eight.

A: Thank you. You’ve got a problem?

2：检查答案，看是否跟你的判断一样，如果不一样，能说出你自己的版本和模板的划分的理由吗？

A: What’s Eve’s number?

B: Four six one eight.

A: Sorry?

B: Four / six / one / eight.

A: That’s not a proper number.

B: Well it has four nine one first, of course.

A: So what’s the full number?

B: Give me strength. / Four / nine / one / four / six / one / eight.

A: Thank you. / You've got a problem?

3：模板的调群切分如上面答案所示，现在打开文件 Task_3.wav，仔细听辩模板是如何实现调群切分和调核的， 有哪些语音策略是模板运用了的？

4：模仿并录下你自己的朗读。对比你对调群切分和调核的语音实现是否和模板相似。如果没有，请反复观察并模仿，直到你自己满意为止

Audacity practice 6

请按照步骤，一一完成！

任务 1；附加疑问句

1：请打开文件 Task_1.wav, 仔细听下列对话，然后在每个对话后面的括号内填上这句附加疑问句的言外之意。（调群和调核已经标注出来）

--He's Czech. --Polish, | isn't he? （ ）

--It's a beautiful day, | isn't it? （ ）

--It looks like rain. --It does, | doesn't it. （ ）

2：现在模仿这几句附加疑问句，直到跟模板相似为止。

任务 2：Leading tones 的应用

1：观察下面已经被切分好的语句，按照这些调群切分，分别录下你的朗读。

Fortunately, / I was wrong.

Then, / I saw a dog.

Today / we're going to do grammar.

Mrs Ashton / will be taking the children.

As for you, / I'll deal with you later.

On the table / you'll find a jug.

If I were you, / I'd wait and see what happens.

2：每句话的第一个调群由于不是完整的句子成分，它们的调核调子都由表非独立意义的降升调呈现。第二个调群都是独立成分，故都是降调。现在打开文件 Task_2.wav，仔细听辩模板是不是这样念的。然后对比你的朗读，如果你自己不是这样实现的，听听哪种版本更好，是你的呢，还是模板的？

3：现在模仿模板的朗读，按照他的方式念出这几句话。所有调群的调核都是非标记性调核，模仿的时候，注意听辩调核的语音实现和非调核部分有何差异。

调核标注：

Fortunately, / I was wrong.

Then, / I saw a dog.

Today / we're going to do grammar.

Mrs Ashton / will be taking the children.

As for you, / I'll deal with you later.

On the table / you'll find a jug.

If I were you, / I'd wait and see what happens.

4：反复听辩并模仿，直到所有成分都跟模板一样。

任务 3：Trailing tones 的应用

1：将下列对话的调群和调核划出来，然后用你觉得合适的语调念出来，并录音。

--What can I do for you, sir?

--I'd like this tie, please.

--And for you, madam?

--Some paper, if you'd be so kind.

2：这两个对话的第二句答语的结构都是主句加非完整结构的附加句，而且主句在前，附加句在后，所以主句都由调核降调完成，而末尾附加句都由升调完成。每句话的调群切分和调核置放如下：

--What can I do for you, sir?

--I'd like this tie, / please.

--And for you, madam?

--Some paper, / if you'd be so kind.

3：打开文件 Task_3.wav，仔细听辩模板是不是按照答案所示调群切分和调核置放实现的这两个对话。再对比你自己的判断，你觉得是模板的比较好，还是你自己的比较好？

4：现在反复模仿模板的实现，注意调核和非调核的语音实现的差异，及调群边界的语音指征。对比模仿直到你自己的实现跟模板相似为止。

任务 4：调头+调核 语调调型实现

1：下面每句话都只有一个调群。调核（下滑线）和调头首重音（`）都已经标记出来。打开文件 Task_4.wav, 仔细听调核和调头的音调走势，将调型填在括号内。

`Better than ever. （调头：　　调核　　）

It's `nearly ready. （调头：　　调核　　）

`Raring to go? （调头：　　调核　　）

Is `that your partner? （调头：　　调核　　）

`Not at the moment. （调头：　　调核　　）

I'm `awfully sorry. （调头：　　调核　　）

`Don't worry. （调头：　　调核　　）

It `doesn't matter. （调头：　　调核　　）

`Saturday's hopeless. （调头：　　调核　　）

We've `only just begun. （调头：　　调核　　）

2：检查答案，看你能否说出这样处理调核调头语调的原因，从情感投入方面思考。

High level + fall: `Better than ever.

It's `nearly ready.

High level + rise: `Raring to go?

Is `that your partner?

High falling + fall-rise: `Not at the moment.

I'm `awfully sorry.

Low level + rise: `Don't worry.

It `doesn't matter.

Low rising + fall (protest): `Saturday's hopeless.

We've `only just begun.

3：现在对照答案，仔细听辩观察模板的实现。模仿并录音，直到跟模板相似为止。

任务 5：低 vs. 高调冠及情感体现

1：打开文件 Task_5.wav，一边听一边看下列低/高调冠对比，看你能否感受其蕴含的情感投入差异 （扫黑部分为调冠）

低调冠	vs.	高调冠
You `mustn't worry.		You `mustn't worry.
I `simply don't believe it.		I `simply don't believe it.
She wasn't.		She wasn't.
I will.		I will.
A handbag?		A handbag?

2：模仿低高调冠的语音实现，对比调核和调头音高走势。直到你自己的朗读跟模板类似为止。

任务 6：对话

1：观察下列对话，在你认为合适的地方划分调群，并标记出调核，调头，调冠，然后用你觉得合适的语调将这段对话念出来，并录音。

Hotel guest: Excuse me, where do I get breakfast?

Receptionist: In the Panorama Restaurant, sir.

Hotel guest: Where's that?

Receptionist: Twenty-seventh floor, sir. Use the lift, over there.

Hotel guest: But the lift only goes to the twenty-fourth floor.

Receptionist: Ah. Use lift number five, sir. That one goes to the twenty-seventh floor.

Hotel guest: I see. Thanks.

2：现在打开文件 Task_6.wav，一边听一边对照答案，看模板是怎么实现这段对话的。

答案：

Hotel guest: Excuse me, / `where do I get breakfast?

Receptionist: In the `Panorama Restaurant, sir.

Hotel guest: `Where's that?

Receptionist: `Twenty-seventh floor, sir. / `Use the lift, / `over there.

Hotel guest: But the `lift only goes to the twenty-fourth floor.

Receptionist: `Ah. `Use lift number five, sir. / That one / `goes to the `twenty-seventh floor.

Hotel guest: I see. / Thanks.

3：对比你自己的朗读，哪些地方跟模板不一样。注意！比较的时候按照下面列举出的顺序来看：

1）调核，调头的位置是否正确；调群切分是否一样？

2）调核的语音实现方式是否一致？

3）调头，调冠的音高走势是否相似？

4）调群边界的语音指征是否相似？

5）调核和非调核是靠那些语音指征区分开来的？

4：比较完后，在按照模板的方式念出这段对话，反复录音直到你自己满意为止。

Appendix VII: Post-test questionnaire

Post-test questionnaire

Welcome!

Thank you so much for your patience with me within these three weeks. This is the last bit of your task!
There will be two parts of questions. Part 1 looks at your self-claimed intonation knowledge that you have learnt from these training sessions. Part 2 aims to elicit your feedback on the training methods you have received. Please answer all the questions based on your best knowledge.

Part I:

1. Which group were you in?
   Audacity
   Praat

2. *Please tick the ONE you think most reflects your understanding of English intonation.*

Statement	Not true at all	Partly true	Very true	I don't know
I know how to manipulate phrasing to make my speech fluent, coherent and disambiguating.				
I can hear how different phrasings change the meaning of a sentence.				
I know how to manipulate nuclear accent to emphasize or contrast the focus/topic of my speech.				
I can hear the focus/topic in other's speech by the placement of nuclear accents.				
I know how to change the tone of my speech to express my intentions.				
I can hear the intentions of a speaker's speech by his/her tones.				
I feel confident about my intonation.				
I now consciously modify my intonation when I speak English.				

Part II:

3. Do you think if you would ever notice these intonation features in native speakers' utterances without the instructions and explanations by the researcher? (Please tick all the answers that apply to you)

   Yes. I can tell most intonation features.
   No. I can't tell any features.
   Partly. I can probably notice something like tones, and be pretty sure of its meanings.
   Partly. I can notice something like tones, but not be sure of its meanings.
   I don't know

4. Please choose ONE option from the multiple choices to each question.

Question	Not useful at all	Not very useful	Quite useful	Extremely useful	Not sure
How useful do you think these weeks of training are for your future self-improvement in English intonation?					
(For Audacity group only) How useful do you think the way that you were asked to 'listen to, identify, and imitate' with the native speakers' models are?					
(For Praat group only) How useful do you think the visual displays of intonation patterns are?					

5. Are you happy about the English models that you were provided with during these training sessions? (You can choose more than one options)

   Not happy
   Quite happy
   More dialects of English would be better
   More real-life materials would be better
   Other (Please specify)

6. *If you used Audacity, your question is:* Do you think it would be more effective if you could 'see' the intonation patterns through a computer?
   If you used Praat, your question is: Do you think it would be equally effective if you couldn't 'see' those intonation patterns?

   Yes
   No
   Probably
   I don't know

7. *If you used Audacity, your question is:* in which aspects you think you can improve more if you "saw" the intonation features?
   If you used Praat, your question is: in which aspects you think you can equally improve even if you didn't "see" the intonation features? (You can tick more options)

   Nuclear tones (pitch)
   Syllable duration
   Vowel precision
   Intensity
   Speaking speed
   Weak forms
   Other (Please specify)

8. Do you think it is worthy of teaching the English intonation system to the students if you were the English teacher? (You can tick more options)

Yes, it is.
No, it isn't.
Yes, it is, but I wouldn't teach it.
Not sure if it is worth teaching
Other (Can you explain why)